

09/884948

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search, using SW model

Run on: September 22, 2003, 12:39:29 ; Search time 44 Seconds
(without alignments)
1194.057 Million cell updates/sec

Title: US-09-884-948-1
Perfect score: 1811
Sequence: 1 DSDRYTPPAEPLDMPDPY.....ITFPRSWNTAPDKVKGWP 331

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 15672573 residues

1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1811	100.0	331	AAV33662
2	1811	100.0	331	AAV33665
3	1811	100.0	331	AAW67770
4	1811	100.0	331	AAW67771
5	1811	100.0	331	AAW67772
6	1811	100.0	332	AAW67771
7	1811	100.0	332	AAW67771
8	1811	100.0	332	AAW67743
9	1811	100.0	346	AAW67743

10	1811	100.0	406	13	AAW22651	Transglutaminase.
11	1811	100.0	407	22	AAW97831	S. mobaraense IPO
12	1811	100.0	407	22	AAW81166	Prepro-transglutamin
13	1811	100.0	407	24	AAW07390	Foreign protein #1
14	1811	100.0	408	15	AAW90049	BTG-contg. sequenc
15	1802	99.5	331	13	AAW22653	Transglutaminase
16	1493	82.4	400	22	AAW47007	Transglutaminase
17	1487	82.1	330	22	AAW81164	Transglutaminase
18	1487	82.1	416	22	AAW97830	S. cinamomeum IPO
19	1487	82.1	416	24	AAW07393	Foreign protein #3
20	1465	80.9	331	17	AAW81570	Transglutaminase
21	1465	80.9	331	23	AAW80745	Streptomyces lydic
22	1432	79.1	330	23	AAW80744	Streptovorticillium
23	1391	76.8	328	20	AAV33666	Streptovorticillium
24	1362	8.9	39	17	AAW81162	Transglutaminase N
25	140	7.7	30	17	AAW81162	peptide fragment o
26	130	7.2	26	17	AAW81174	peptide fragment o
27	124	6.8	25	17	AAW81175	peptide fragment o
28	119	6.6	23	17	AAW81176	peptide fragment o
29	116	6.4	23	17	AAW81176	peptide fragment o
30	114.5	6.3	889	22	AAW65872	Drosophila melanog
31	114.5	6.3	1217	24	AAW60377	Bombyx mori serici
32	112.5	6.2	1161	22	AAW70449	Drosophila melanog
33	110	6.1	736	22	AAW59845	AAV3B capsid prote
34	109	6.0	21	17	AAW82022	peptide fragment o
35	108	6.0	558	23	AAW805719	Human testis deriv
36	108	6.0	1005	21	AAW43890	Human cancer assoc
37	108	6.0	1382	18	AAW11687	Human metastas-s-a
38	107	5.9	736	22	AAW59846	AAV3B capsid prote
39	106.5	5.9	736	21	AAW71167	Adeno-associated v
40	106.5	5.9	736	22	AAW59847	AAV6 capsid protei
41	106	5.9	23	17	AAW81168	peptide fragment o
42	105	5.8	764	21	AAW53024	Arabidopsis thalia
43	105	5.8	905	21	AAW53023	Arabidopsis thalia
44	105	5.8	919	21	AAW53022	Arabidopsis thalia
45	104	5.7	24	17	AAW81169	peptide fragment o

ALIGNMENTS

RESULT 1
AAV33662 standard; Protein; 331 AA.
ID AAV33662; AA.
AC AAV33662;
XX
XX
XX 07-JAN-2000 (first entry)
XX
XX
XX Streptovorticillium S-8112 transglutaminase protein fragment.
DE
XX
XX Transglutaminase; food industry; pharmaceutical industry; texture;
KW Cosmetic industry; proteinaceous material; gel strength; viscosity;
KW breaking strength; elasticity; taste.
XX
XX Streptovorticillium sp.
OS
XX
XX
XX DE19814860-A1.
XX
XX
XX 07-OCT-1999.
XX
XX
XX 02-APR-1998; 98DE-1014860.
XX
XX
XX 02-APR-1998; 98DE-1014860.
XX
XX
XX 02-APR-1998; 98DE-1014860.
XX
XX
XX (FUCH) FUCHSBAUER H.
XX
XX Fuchsbaue H. Pasternack R. Dorsch S. Otterbach J. Roberek I.
XX
XX Mainusch M. Dauscher C.
XX
XX WPI: 1999-552288/47.
XX
XX N-PsDB; AA223653.
XX

PT Bacterial transglutaminase polypeptides useful for polymerizing
 PT proteins, e.g. to modify the properties of food, pharmaceutical or
 PT cosmetic products -
 XX
 PS Claim 1; Page 23-24; 44pp; German.
 XX
 CC This invention describes a novel bacterial transglutaminase polypeptide.
 CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic
 CC industries, to polymerize proteinaceous materials in order to improve
 CC their properties, e.g. texture, gel strength, breaking strength,
 CC viscosity, elasticity or taste. (I) can also be used to immobilize
 CC enzymes and antibodies. This sequence represents a transglutaminase
 CC isolated from *Streptococcus* sp. S-8112.
 CC
 XX Sequence 331 AA;
 SQ
 Query Match 100.0%; Score 1811; DB 20; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOYVSHRDKKQMTBEGREMT 60
 DB 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOYVSHRDKKQMTBEGREMT 60
 QY 61 SYGCVGVTVNWSGQYPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
 DB 61 SYGCVGVTVNWSGQYPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
 QY 121 KGFQRAEVASVMMRLAENADESAVLDNLKEELANGNDALRNEDARPPYSALRNTPSF 180
 DB 121 KGFQRAEVASVMMRLAENADESAVLDNLKEELANGNDALRNEDARPPYSALRNTPSF 180
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDENI 240
 DB 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDENI 240
 QY 241 PRSPTSGEGFVNPDYGFAGQTEADADKTVTHGNHYAPKGSIGAHVYESKFRWMS 300
 DB 241 PRSPTSGEGFVNPDYGFAGQTEADADKTVTHGNHYAPKGSIGAHVYESKFRWMS 300
 QY 301 GYSDPDRGAVVITFIPKSWNTAPDKVKQGW 331
 DB 301 GYSDPDRGAVVITFIPKSWNTAPDKVKQGW 331
 RESULT 2
 ID AAY33665 standard; Protein; 331 AA.
 AC AAY33665;
 XX
 DT 07-JAN-2000 (first entry)
 XX
 DE Streptococcus mobaraense transglutaminase protein fragment.
 XX
 KM Transglutaminase; food industry; pharmaceutical industry; texture;
 KM cosmetic industry; proteinaceous material; gel strength; viscosity;
 KM breaking strength; elasticity; taste.
 XX
 OS Streptococcus mobaraense.
 XX
 PN DE19814860-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 02-APR-1998; 98DE-1014860.
 XX
 PR 02-APR-1998; 98DE-1014860.
 XX
 PA (FUCHS) FUCHSBAUER H.
 PI Fuchsbauer H., Pasternack R., Dorsch S., Otterbach J., Robenek I;
 PI Mainusch M., Dauscher C;

XX
 DR WP1; 1999-552288/47.
 XX
 PT Bacterial transglutaminase polypeptides useful for polymerizing
 PT proteins, e.g. to modify the properties of food, pharmaceutical or
 PT cosmetic products -
 XX
 PS Claim 12; Page 25-26; 44pp; German.
 XX
 CC This invention describes a novel bacterial transglutaminase polypeptide.
 CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic
 CC industries, to polymerize proteinaceous materials in order to improve
 CC their properties, e.g. texture, gel strength, breaking strength,
 CC viscosity, elasticity or taste. (I) can also be used to immobilize
 CC enzymes and antibodies. This sequence represents a transglutaminase
 CC protein fragment isolated from *Streptococcus* sp. S-8112.
 CC
 XX Sequence 331 AA;
 SQ
 Query Match 100.0%; Score 1811; DB 20; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOYVSHRDKKQMTBEGREMT 60
 DB 1 DSDRVTTPAEPPLDMPDPYRPSYGRATVNNYIRKQOYVSHRDKKQMTBEGREMT 60
 QY 61 SYGCVGVTVNWSGQYPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
 DB 61 SYGCVGVTVNWSGQYPTNRLAFASFDREPKNELKGRPRSGETPAEFGVAKESFDEE 120
 QY 121 KGFQRAEVASVMMRLAENADESAVLDNLKEELANGNDALRNEDARPPYSALRNTPSF 180
 DB 121 KGFQRAEVASVMMRLAENADESAVLDNLKEELANGNDALRNEDARPPYSALRNTPSF 180
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDENI 240
 DB 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDENI 240
 QY 241 PRSPTSGEGFVNPDYGFAGQTEADADKTVTHGNHYAPKGSIGAHVYESKFRWMS 300
 DB 241 PRSPTSGEGFVNPDYGFAGQTEADADKTVTHGNHYAPKGSIGAHVYESKFRWMS 300
 QY 301 GYSDPDRGAVVITFIPKSWNTAPDKVKQGW 331
 DB 301 GYSDPDRGAVVITFIPKSWNTAPDKVKQGW 331
 RESULT 3
 ID AAM67770 standard; Protein; 331 AA.
 AC AAM67770;
 XX
 DT 01-APR-1999 (first entry)
 XX
 DE A transglutaminase enzyme sequence.
 XX
 KM Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;
 KM cosmetic; meat quality; microcapsule production; high thermal stability;
 KM carrier; immobilised enzyme.
 XX
 OS Streptococcus sp.
 XX
 PN EP889133-A2.
 XX
 PD 07-JAN-1999.
 XX
 PF 02-JUL-1998; 98EP-0112315.
 XX
 PR 04-JUL-1997; 97JP-0180010.
 XX
 PA (AJIN) AJINOMOTO CO INC.

XX Miwa T, Nakamura N, Seguro K, Yokoyama K;
 XX WPI; 1999-062664/06.
 DR N-PSDB; AAV81507.
 XX
 XX New microbial transglutaminase with N-terminal aspartic acid deleted
 PT - allowing high level recombinant production without added
 PT methionine in *E. coli*, useful in production of gelled foods,
 PT cosmetics etc.
 XX
 XX Claim 1; Page 12-14; 56pp; English.

CC The present sequence represents a transglutaminase of *Streptococcus*
 CC *sp.*. The specification describes a new microbial transglutaminase that
 CC has the N-terminal aspartic acid of the present transglutaminase
 CC deleted. Eliminating the N-terminal Asp from microbial transglutaminase
 CC allows efficient removal of the terminal methionine residue added when
 CC the protein is expressed in *Escherichia coli*. The *E. coli* methionine
 CC aminopeptidase acts well on Met-Ser but only poorly on Met-Asp, so
 CC problems of antigenicity associated with Met-terminated proteins are
 CC avoided. Recombinant transglutaminase is used to produce gelled foods
 CC (jellies, yogurt and cheeses) or cosmetics, to improve the quality of
 CC meat, in the production of materials for microcapsules of high thermal
 CC stability and as a carrier for immobilised enzymes.
 CC
 XX Sequence 331 AA:

Query Match 100.0%; Score 1811; DB 20; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRVTPEAPEDLMDPDPYPSYGRATVNNYIRKMOQVYSHRDKQOMTEBOREWL 60
 DB 1 DSDRVTPEAPEDLMDPDPYPSYGRATVNNYIRKMOQVYSHRDKQOMTEBOREWL 60
 QY 61 SYGCVGVTVNWSGQYPTNRLAFASFDEDFKKNELKNGPRSGETRAEFGRAVAKESFDEE 120
 DB 61 SYGCVGVTVNWSGQYPTNRLAFASFDEDFKKNELKNGPRSGETRAEFGRAVAKESFDEE 120
 QY 121 KGQORAREVASVNNRLENAHDESAVLDNLKKELANGNDALNEDARSFPYSALRNTPSF 180
 DB 121 KGQORAREVASVNNRLENAHDESAVLDNLKKELANGNDALNEDARSFPYSALRNTPSF 180
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPAPFPAGTGLVDMGRDNT 240
 DB 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPAPFPAGTGLVDMGRDNT 240
 QY 241 PRSPTPSGEGFVNFYGFAGQTEADADKTWTHGNHHAHNGSLGAMHYESKFRMWS 300
 DB 241 PRSPTPSGEGFVNFYGFAGQTEADADKTWTHGNHHAHNGSLGAMHYESKFRMWS 300
 QY 301 GYSDPDRGAYVITFIKSNMTAPDKVYKQGNP 331
 DB 301 GYSDPDRGAYVITFIKSNMTAPDKVYKQGNP 331

RESULT 4
 AAB81161
 ID AAB81161 standard; protein; 331 AA.

XX AAB81161;
 AC
 XX 13-JUN-2001 (first entry)
 DT
 XX Transglutaminase related protein SEQ ID 5.
 DE
 XX
 XX Coryneform bacteria; transglutaminase; food processing.
 KM
 XX Streptococcus mobaraense.
 DS
 XX WO200123591-A1.
 XX

PD 05-APR-2001.
 XX
 XX 29-SEP-2000; 2000WO-JP06780.
 PF
 XX 30-SEP-1999; 99JP-0280098.
 PR
 XX 28-JUN-2000; 2000JP-0194043.
 PR
 XX (AJIN) AJINOMOTO CO INC.
 PA
 XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Mateui H;
 PI
 XX WPI; 2001-266172/27.
 DR

PT Efficient secretory production of foreign proteins e.g.
 PT transglutaminase employing transformant coryneform bacterium, simply on
 PT industrial scale with direct recovery for use in food processing and
 PT pharmaceutical industry -
 XX
 XX Claim 22; Page 79-81; 151pp; Japanese.

CC This invention relates to a process for the production of a foreign
 CC secretory protein through the construction of a recombinant coryneform
 CC bacterium. The coryneform bacterium is transformed with an expression
 CC construct in which DNA encoding a target foreign protein pro-structure is
 CC ligated to the downstream region of DNA encoding the signal peptide
 CC domain of a coryneform bacterial protein. Following transformation with
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from
 CC the expressed protein. Transglutaminases produced using this process are
 CC useful in the food processing and pharmaceutical industries. The present
 CC sequence represents a transglutaminase related protein, which can be
 CC used in the method of the invention.
 CC
 XX Sequence 331 AA;

Query Match 100.0%; Score 1811; DB 22; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRVTPEAPEDLMDPDPYPSYGRATVNNYIRKMOQVYSHRDKQOMTEBOREWL 60
 DB 1 DSDRVTPEAPEDLMDPDPYPSYGRATVNNYIRKMOQVYSHRDKQOMTEBOREWL 60
 QY 61 SYGCVGVTVNWSGQYPTNRLAFASFDEDFKKNELKNGPRSGETRAEFGRAVAKESFDEE 120
 DB 61 SYGCVGVTVNWSGQYPTNRLAFASFDEDFKKNELKNGPRSGETRAEFGRAVAKESFDEE 120
 QY 121 KGQORAREVASVNNRLENAHDESAVLDNLKKELANGNDALNEDARSFPYSALRNTPSF 180
 DB 121 KGQORAREVASVNNRLENAHDESAVLDNLKKELANGNDALNEDARSFPYSALRNTPSF 180
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPAPFPAGTGLVDMGRDNT 240
 DB 181 KERNGNHDPSSRMKAVIYSKHFWSGODRSSADKRYGDPAPFPAGTGLVDMGRDNT 240
 QY 241 PRSPTPSGEGFVNFYGFAGQTEADADKTWTHGNHHAHNGSLGAMHYESKFRMWS 300
 DB 241 PRSPTPSGEGFVNFYGFAGQTEADADKTWTHGNHHAHNGSLGAMHYESKFRMWS 300
 QY 301 GYSDPDRGAYVITFIKSNMTAPDKVYKQGNP 331
 DB 301 GYSDPDRGAYVITFIKSNMTAPDKVYKQGNP 331

RESULT 5
 ABB06742
 ID ABB06742 standard; Protein; 331 AA.

XX ABB06742;
 AC
 XX 12-JUN-2002 (first entry)
 DT
 XX Streptococcus mobaraense transglutaminase protein SEQ ID NO:2.
 DE
 XX

KM Microbial; transglutaminase; protein co-ordinate data; stereo-structure;
 KW X-ray analysis; crystalline structure; enzyme; food processing;
 KW thermal stability.
 XX Streptomyces mobaraensis.
 OS
 XX MO200214518-A1.
 XX
 XX 21-FEB-2002.
 XX
 XX 15-AUG-2001; 2001MO-JP07038.
 XX
 XX 17-AUG-2000; 2000JP-0247664.
 XX
 XX 27-DEC-2000; 2000JP-0396695.
 XX
 XX (AJIN) AJINOMOTO CO INC.
 PA
 PI Kashiwagi T, Shimba N, Ishikawa K, Suzuki E, Yokoyama K,
 PI Hirayama K;
 XX
 XX WPI: 2002-269198/31.
 DR N-PSDB; ABL50235.
 PS
 PS Claim 2; Page 100-102; 126pp; Japanese.
 XX
 XX The present invention describes a method for designing and constructing
 CC a variant transglutaminase by estimating the binding site of
 CC Streptococcus thermophilus-derived transglutaminase (MTG) to the
 CC substrate based on the stereo-structure obtained by X-ray analysis of
 CC the crystalline structure of MTG crystals, and e.g. substituting amino
 CC acid residues located at the substrate-binding site of the
 CC transglutaminase. The method can be used for designing and constructing
 CC a variant transglutaminase. The obtained transglutaminases can be used
 CC in food processing. The modified transglutaminases have improved
 CC transglutaminase activity and thermal stability, substrate-specificity
 CC and an less required optimum pH. The present sequence represents a
 CC transglutaminase isolated from Streptococcus thermophilus
 CC (also called Streptomyces mobaraensis).
 CC
 XX
 XX
 XX Sequence 331 AA;
 SQ
 SQ Query Match 100.0%; Score 1811; DB 23; Length 331;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLRMDPPRPSYGAETVANNYTRKQOYVSHRDRKQOMTEBQREML 60
 DB 1 DSDDRVTPPAEPLRMDPPRPSYGAETVANNYTRKQOYVSHRDRKQOMTEBQREML 60
 QY 61 SYGCVGVTVWNSGQYPTNRLAFASFEDEFRFXNELKNGRPRSGETRAFFBGRVAKESFDEE 120
 DB 61 SYGCVGVTVWNSGQYPTNRLAFASFEDEFRFXNELKNGRPRSGETRAFFBGRVAKESFDEE 120
 QY 121 KGFORAREVAVNMRALENAHDESAVLTNKKELANGNDALRNNDASPPYSALRNTSPF 180
 DB 121 KGFORAREVAVNMRALENAHDESAVLTNKKELANGNDALRNNDASPPYSALRNTSPF 180
 QY 181 KERNGNNDHPSRMKAVIVSKLFWNSGQDRSSADKRYADPDAPFAPGTGLVDMSTRNI 240
 DB 181 KERNGNNDHPSRMKAVIVSKLFWNSGQDRSSADKRYADPDAPFAPGTGLVDMSTRNI 240
 QY 241 PRSPFSPGSGFVNPYGMFGAGCTEADADCTWTGNNHYHAHNSGLAMNHYBSKPRANSE 300
 DB 241 PRSPFSPGSGFVNPYGMFGAGCTEADADCTWTGNNHYHAHNSGLAMNHYBSKPRANSE 300
 QY 301 GYSDPDRGAVYITFLPKSWNTAPDKVKGQMP 331
 DB 301 GYSDPDRGAVYITFLPKSWNTAPDKVKGQMP 331

RESULT 6
 ID AAM67771 standard; Protein; 332 AA.
 AAM67771
 AC AAM67771;
 DT 01-APR-1999 (first entry)
 XX
 DE Protein encoded by high expression transglutaminase gene.
 KW Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;
 KW cosmetic; meat quality; microcapsule production; high thermal stability;
 KW carrier; immobilised enzyme.
 XX
 XX Synthetic.
 OS Streptococcus thermophilus sp.
 XX
 XX EP889133-A2.
 XX
 XX 07-JAN-1999.
 PD
 XX
 XX 02-JUL-1998; 98EP-0112315.
 PF
 XX
 XX 04-JUL-1997; 97JP-0180010.
 PR
 XX
 XX (AJIN) AJINOMOTO CO INC.
 PA
 PI Miwa T, Nakamura N, Seguro K, Yokoyama K;
 PI N-PSDB; AAV81508.
 XX
 XX WPI: 1999-062664/06.
 DR N-PSDB; AAV81508.
 XX
 XX New microbial transglutaminase with N-terminal aspartic acid deleted
 PT - allowing high level recombinant production without added
 PT methionine in E. coli, useful in production of gelled foods,
 PT cosmetics etc.
 PT
 PS
 PS Example 1; Page 18-23; 56pp; English.
 XX
 XX The present sequence is encoded by the high expression transglutaminase
 CC gene present in plasmid pTRMTG-02. The gene is derived from
 CC Streptococcus thermophilus sp., and is codon altered, using oligonucleotides
 CC AAV81521-60, for expression in Escherichia coli. The specification
 CC describes a new microbial transglutaminase that has the N-terminal
 CC aspartic acid of transglutaminase deleted. Eliminating the N-terminal
 CC Asp from microbial transglutaminase allows efficient removal of the
 CC terminal Met residue added when the protein is expressed in E. coli.
 CC The E. coli methionine aminopeptidase acts well on Met-Ser but only
 CC poorly on Met-Asp, so problems of antigenicity associated with
 CC Met-terminated proteins are avoided. Recombinant transglutaminase is
 CC used to produce gelled foods (jellies, yogurt and cheeses) or cosmetics,
 CC to improve the quality of meat, in the production of materials for
 CC microcapsules of high thermal stability and as a carrier for immobilised
 CC enzymes.
 CC
 XX
 XX Sequence 332 AA;
 SQ
 SQ Query Match 100.0%; Score 1811; DB 20; Length 332;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLRMDPPRPSYGAETVANNYTRKQOYVSHRDRKQOMTEBQREML 60
 DB 2 DSDDRVTPPAEPLRMDPPRPSYGAETVANNYTRKQOYVSHRDRKQOMTEBQREML 61
 QY 61 SYGCVGVTVWNSGQYPTNRLAFASFEDEFRFXNELKNGRPRSGETRAFFBGRVAKESFDEE 120
 DB 62 SYGCVGVTVWNSGQYPTNRLAFASFEDEFRFXNELKNGRPRSGETRAFFBGRVAKESFDEE 121
 QY 121 KGFORAREVAVNMRALENAHDESAVLTNKKELANGNDALRNNDASPPYSALRNTSPF 180
 DB 121 KGFORAREVAVNMRALENAHDESAVLTNKKELANGNDALRNNDASPPYSALRNTSPF 180

Db 122 KGFQARREAVSWANRLAENAHDESAVLDNLKKELANGNALNEDAPSPFYSAIRNTSPF 181
 QY 181 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFPAPGTGLVDMASDRNT 240
 Db 182 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFPAPGTGLVDMASDRNT 241
 QY 241 PRSPFSGEGFVNPFDYGMFGAQTADAKTYWTHGNHHAHPNGLAMHYESKFRWMS 300
 Db 242 PRSPFSGEGFVNPFDYGMFGAQTADAKTYWTHGNHHAHPNGLAMHYESKFRWMS 301
 QY 301 GYSDPDRGAVITFIPKSWNTAPDKVKQGW 331
 Db 302 GYSDPDRGAVITFIPKSWNTAPDKVKQGW 332

RESULT 7
 ID AAB12809 standard; Protein; 332 AA.
 AC AAB12809;
 DT 24-NOV-2000 (first entry)
 DE Transglutaminase protein sequence SEQ ID NO:1.
 KM Transglutaminase; gelled food; jelly; yoghurt; gelled cosmetic;
 KM cheese.
 OS Unidentified.
 PN WO20040706-A1.
 PD 13-JUL-2000.
 PF 24-DEC-1999; 99NO-JP07250.
 PR 28-DEC-1998; 98JP-0373131.
 PA (AJIN) AJINOMOTO CO INC.
 PI Yokoyama K, Ono K, Ejima D;
 DR WPI; 2000-475826/41.
 DR N-PSDB; AAA73025.
 PT Production of active transglutaminase from denatured enzyme by
 PT two-stage refolding process for industrial production of active enzyme
 PT for use in food production
 PS Disclosure; Page 48-50; 74pp; Japanese.
 CC The present invention describes a method for producing active
 CC transglutaminase from denatured enzyme. The method comprises: (i) forming
 CC an intermediate structure of the enzyme having transglutaminase activity
 CC under acidic conditions in an aqueous medium; and (ii) forming a higher-
 CC level structure of the enzyme having transglutaminase activity under
 CC neutral conditions in an aqueous medium. The method can be used for
 CC industrial production of active transglutaminase from denatured material
 CC (such as recombinant transglutaminase) which can be used in the food
 CC industry for the production of gelled foods such as jellies, yoghurts
 CC and cheeses, and for the production of gelled cosmetics. The present
 CC sequence represents a transglutaminase which is used in the
 CC exemplification from the present invention.

Sequence 332 AA;
 SQ

Query Match 100.0%; Score 1811; DB 21; Length 332;
 Best Local Similarity 100.0%; Pred. No. 2, 9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRVTTPAPPIJDRMPDPYRPSYGRATVNNYIRKQOYVSHDGRKQOMTEOREWL 60
 2 DSDRVTTPAPPIJDRMPDPYRPSYGRATVNNYIRKQOYVSHDGRKQOMTEOREWL 61

QY 61 SYGCVGTWNSQYPTNRLAPAFEDSDREKRLKNGRPSGETTAEEGRVANAESFDE 120
 Db 62 SYGCVGTWNSQYPTNRLAPAFEDSDREKRLKNGRPSGETTAEEGRVANAESFDE 121
 QY 121 KGFQARREAVSWANRLAENAHDESAVLDNLKKELANGNALNEDAPSPFYSAIRNTSPF 180
 Db 122 KGFQARREAVSWANRLAENAHDESAVLDNLKKELANGNALNEDAPSPFYSAIRNTSPF 181
 QY 181 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFPAPGTGLVDMASDRNT 240
 Db 182 KERNGNHDPSSMKAVIYSKHPWSGODRSSADKRYGDPDAFPAPGTGLVDMASDRNT 241
 QY 241 PRSPFSGEGFVNPFDYGMFGAQTADAKTYWTHGNHHAHPNGLAMHYESKFRWMS 300
 Db 242 PRSPFSGEGFVNPFDYGMFGAQTADAKTYWTHGNHHAHPNGLAMHYESKFRWMS 301
 QY 301 GYSDPDRGAVITFIPKSWNTAPDKVKQGW 331
 Db 302 GYSDPDRGAVITFIPKSWNTAPDKVKQGW 332

RESULT 8
 ID ABB06743 standard; Protein; 332 AA.
 AC ABB06743;
 DT 12-JUN-2002 (first entry)
 DE Streptococcus thermophilus molarase transglutaminase protein SEQ ID NO:4.
 KM Microbial; transglutaminase; protein co-ordinate data; stereo-structure;
 KM X-ray analysis; crystalline structure; enzyme; food processing;
 KM thermal stability.
 OS Streptococcus molarase.
 PN WO200214518-A1.
 PD 21-FEB-2002.
 PF 15-AUG-2001; 2001MO-JP07038.
 PR 17-AUG-2000; 2000JP-0247664.
 PR 27-DEC-2000; 2000JP-0396695.
 PA (AJIN) AJINOMOTO CO INC.
 PI Kashiwagi T, Shima N, Ishikawa K, Suzuki E, Yokoyama K;
 PI Hileyama K;
 DR WPI; 2002-269198/31.
 DR N-PSDB; ABL50236.
 PT Designing and constructing a variant of Streptococcus thermophilus
 PT molarase-originated transglutaminase (MTG), for use in food
 PT processing, comprises estimating the binding site of MTG based on its
 PT stereo-structure
 PS Example 4; Page 105-106; 126pp; Japanese.
 CC The present invention describes a method for designing and constructing
 CC a variant transglutaminase by estimating the binding site of
 CC Streptococcus thermophilus molarase-originated transglutaminase (MTG) to the
 CC substrate based on the stereo-structure obtained by X-ray analysis of
 CC the crystalline structure of MTG crystals, and e.g. substituting amino

CC acid residues located at the substrate-binding site of the
 CC transglutaminase. The method can be used for designing and constructing
 CC a variant transglutaminase. The obtained transglutaminases can be used
 CC in food processing. The modified transglutaminases have improved
 CC transglutaminase activity and thermal stability, substrate-specificity
 CC and an less required optimum pH. The present sequence represents a
 CC transglutaminase isolated from *Streptococcus thermophilus* mobarensis
 CC (also called *Streptococcus mobarensis*).
 XX
 SQ Sequence 332 AA;
 Query Match 100.0%; Score 1811; DB 23; Length 332;
 Best Local Similarity 100.0%; Pred. No. 2.9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDMDPPYRPSYGRATVNNYIRKQVYSHRDGRKQWTEQREWL 60
 DB 2 DSDDRVTPPAEPLDMDPPYRPSYGRATVNNYIRKQVYSHRDGRKQWTEQREWL 61
 QY 61 SYGCVGVTWNSGQYPTNRLAFASPEDERFKNELANGRPSCGTRAEFGRAVAKSPDEE 120
 DB 62 SYGCVGVTWNSGQYPTNRLAFASPEDERFKNELANGRPSCGTRAEFGRAVAKSPDEE 121
 QY 121 KGFQARREVASVNNRALLENAHDESAYLDNLKKELANGNDALRNEDARSFFYSALRNTPSF 180
 DB 122 KGFQARREVASVNNRALLENAHDESAYLDNLKKELANGNDALRNEDARSFFYSALRNTPSF 181
 QY 181 KERNGNHNDSRRKAAVYISKHFWSGQDSSADRRKXGDPDARPPAPGTGLVDSRDNI 240
 DB 182 KERNGNHNDSRRKAAVYISKHFWSGQDSSADRRKXGDPDARPPAPGTGLVDSRDNI 241
 QY 241 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMHVESKFRMNS 300
 DB 242 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMHVESKFRMNS 301
 QY 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331
 DB 302 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 332

RESULT 9
 AAR49048
 ID AAR49048 standard; Protein; 346 AA.
 XX
 AC AAR49048;
 XX
 DT 10-MAR-2003 (updated)
 DT 20-SEP-1994 (first entry)
 XX
 DE Bacterial transglutaminase.
 XX
 KM Bacterial transglutaminase; BTG; expression;
 KM active; inactive; inclusion body.
 XX
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN JF06030771-A.
 XX
 PD 08-FEB-1994.
 XX
 PF 14-JUL-1992; 92JP-0187038.
 XX
 PR 14-JUL-1992; 92JP-0187038.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI, 1994-079294/10.
 DR N-PSDB; AAQ55983.
 XX
 PT Prepn. of bacterial trans-glutaminase in large ants. - by
 PT expression of fusion protein in *E. coli* bacterial
 PT trans-glutaminase

XX
 PS Claim 1; Page 7-8; 13pp; Japanese.
 XX
 CC A fused protein contains amino acids 16-346 of BTG (AAQ55983)
 CC and a hydrophilic peptide at the amino terminal.
 CC Expression of DNA encoding this protein in *E. coli* allow large
 CC scale prodn. of BTG. An active BTG can be prepd. from the
 CC inactive fused protein inclusion body.
 CC (Updated on 10-MAR-2003 to add missing OS field.)
 XX
 SQ Sequence 346 AA;
 Query Match 100.0%; Score 1811; DB 15; Length 346;
 Best Local Similarity 100.0%; Pred. No. 3.1e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDMDPPYRPSYGRATVNNYIRKQVYSHRDGRKQWTEQREWL 60
 DB 16 DSDDRVTPPAEPLDMDPPYRPSYGRATVNNYIRKQVYSHRDGRKQWTEQREWL 75
 QY 61 SYGCVGVTWNSGQYPTNRLAFASPEDERFKNELANGRPSCGTRAEFGRAVAKSPDEE 120
 DB 76 SYGCVGVTWNSGQYPTNRLAFASPEDERFKNELANGRPSCGTRAEFGRAVAKSPDEE 135
 QY 121 KGFQARREVASVNNRALLENAHDESAYLDNLKKELANGNDALRNEDARSFFYSALRNTPSF 180
 DB 136 KGFQARREVASVNNRALLENAHDESAYLDNLKKELANGNDALRNEDARSFFYSALRNTPSF 195
 QY 181 KERNGNHNDSRRKAAVYISKHFWSGQDSSADRRKXGDPDARPPAPGTGLVDSRDNI 240
 DB 196 KERNGNHNDSRRKAAVYISKHFWSGQDSSADRRKXGDPDARPPAPGTGLVDSRDNI 255
 QY 241 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMHVESKFRMNS 300
 DB 256 PRSPFSPGEGFVNFDYGMFGAQTEDADKVTWTHGNHYHAPNGSLGAMHVESKFRMNS 315
 QY 301 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 331
 DB 316 GYSDFDRGAYVITFIKSWNTAPDKVKQGW 346

RESULT 10
 AAR22651
 ID AAR22651 standard; Protein; 406 AA.
 XX
 AC AAR22651;
 XX
 DT 25-MAR-2003 (updated)
 DT 09-OCT-1992 (first entry)
 XX
 DE Transglutaminase.
 XX
 KM BTG; acyl rearrangement; deamination.
 XX
 OS Actinomycetes and *Streptococcus*.
 OS Key
 FH Peptide Location/Qualifiers
 FT 1..75
 FT /label= sig_peptide
 FT 76..406
 FT /label= mat_BTG
 XX
 EP481504-A.
 XX
 PD 22-APR-1992.
 XX
 PF 18-OCT-1991; 91EP-0117813.
 XX
 PR 19-OCT-1990; 90JP-0282566.
 XX
 PA (AMAN) AMANO PHARM KK.
 XX
 PT (AJIN) AJINOMOTO KK.

PI Ando K, Arafuka S, Koikeda S, Matsui H, Takagi H, Washizu K;
 XX WPI; 1992-133808/17.
 DR N-PSDB; AAQ24197, AAQ24201.
 XX
 PT DNA fragment encoding trans:glutaminase - is inserted into
 XX vector, e.g. Pml1053-BTG, for protein expression
 XX
 PS Disclosure; Page 42 and 44; 55 pp; English.
 CC The mature transglutaminase enzyme (BTG) can be derived from two
 CC different species, Streptococcus lactis and Actinomyces. It
 CC catalyzes an acyl rearrangement reaction of a gamma-carboxamide gp.
 CC of glutamine. It introduces intra- or intermolecular formation of
 CC epsilon-(gamma-Gln)-Lys cross-linking when an epsilon-amino gp. of
 CC a Lys residue acts as an acyl receptor. When water acts as an acyl
 CC acceptor the enzyme accelerates the conversion of Gln residues to Gln
 CC residues by deamination.
 CC The enzyme is used in the prodn. of gelled foods, gelled cosmetics,
 CC yogurt, gelatin, cheese etc. It is also used in the prodn. of
 CC thermally stable materials such as microcapsules and carriers of
 CC immobilized enzymes. The DNA sequence given allows the prodn. of
 CC BTG efficiently and in large quantity.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX SQ Sequence 406 AA;
 Query Match 100.0%; Score 1811; DB 13; Length 406;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDMDPPRPSYGRAGETVNNYIRKQVYSHRDGRKQMTTEOREWL 60
 DB DSDDRVTPPAEPLDMDPPRPSYGRAGETVNNYIRKQVYSHRDGRKQMTTEOREWL 135
 QY 61 SYGCVGVTVNWSGQYPTNRLAFASPDDEFKNEIKNGRPSGERTAFEGRYAKESFDEE 120
 DB SYGCVGVTVNWSGQYPTNRLAFASPDDEFKNEIKNGRPSGERTAFEGRYAKESFDEE 195
 QY 121 KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 180
 DB KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 255
 QY 136 KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 255
 DB KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 315
 QY 181 KERNGNHDPBRMKAIVYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSDRNI 240
 DB KERNGNHDPBRMKAIVYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSDRNI 315
 QY 241 PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 300
 DB PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 375
 QY 316 PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 375
 DB PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 375
 QY 301 GYSDFDRGAYVITFIPKSNMTAPDKVKQGP 331
 DB GYSDFDRGAYVITFIPKSNMTAPDKVKQGP 406
 RESULT 11
 AAB97831
 ID AAB97831 standard; Protein; 407 AA.
 XX
 AC AAB97831;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE S. mobaraense IPO 13819 transglutaminase protein SEQ ID NO.4.
 XX
 KM Streptococcus lactis cinamomum IPO 12852; Streptomyces; actinomycete;
 XX Streptococcus lactis mobaraense IPO 13819; transglutaminase.
 XX
 OS Streptococcus lactis mobaraensis.
 XX
 PN WO200129187-A1.
 XX

PD 26-APR-2001.
 XX
 PF 13-OCT-2000; 2000MO-JP07135.
 XX
 PR 18-OCT-1999; 99JP-0295649.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 PI Taguchi S, Komose H;
 XX
 DR WPI; 2001-300330/31.
 DR N-PSDB; AAH20188.
 XX
 PT Streptomyces sp. carrying an actinomycete-derived gene and promoter for
 XX producing high yields of transglutaminase
 PS Disclosure; Page 33-36; 41pp; Japanese.
 CC The present invention describes a Streptomyces sp. containing a gene
 CC construct comprising actinomycete-derived transglutaminase gene and
 CC promoter. Also described are methods for producing pro-transglutaminase
 CC and active transglutaminase. The gene construct can be used in the
 CC production of large amounts of transglutaminase. The present sequence
 CC represents Streptococcus lactis mobaraense IPO 13819 transglutaminase,
 CC which is given in the exemplification of the present invention.
 CC
 XX SQ Sequence 407 AA;
 Query Match 100.0%; Score 1811; DB 22; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDMDPPRPSYGRAGETVNNYIRKQVYSHRDGRKQMTTEOREWL 60
 DB DSDDRVTPPAEPLDMDPPRPSYGRAGETVNNYIRKQVYSHRDGRKQMTTEOREWL 136
 QY 61 SYGCVGVTVNWSGQYPTNRLAFASPDDEFKNEIKNGRPSGERTAFEGRYAKESFDEE 120
 DB SYGCVGVTVNWSGQYPTNRLAFASPDDEFKNEIKNGRPSGERTAFEGRYAKESFDEE 196
 QY 121 KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 180
 DB KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 256
 QY 197 KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 256
 DB KGFQRAEYASVWNRALNADHDSAYLDNLKELANGDALARNDARSFFYSALRNTPSF 316
 QY 181 KERNGNHDPBRMKAIVYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSDRNI 240
 DB KERNGNHDPBRMKAIVYSKHPWSGQDRSSADKRYGDPDAPRPAOTGLVMSDRNI 316
 QY 241 PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 300
 DB PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 376
 QY 317 PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 376
 DB PRSPSPGEGFVNDYGMFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFRNMSE 376
 QY 301 GYSDFDRGAYVITFIPKSNMTAPDKVKQGP 331
 DB GYSDFDRGAYVITFIPKSNMTAPDKVKQGP 407
 RESULT 12
 AAB81166
 ID AAB81166 standard; Protein; 407 AA.
 XX
 AC AAB81166;
 XX
 DT 13-JUL-2001 (first entry)
 XX
 DE Prepro-transglutaminase amino acid sequence.
 XX
 KM Corynebacterium bacteria; transglutaminase; food processing.
 XX Streptococcus lactis mobaraense.
 XX
 OS Streptococcus lactis mobaraensis.
 XX
 PN WO200123591-A1.
 XX

PD 05-APR-2001.
 XX
 XX 29-SEP-2000; 2000MO-JP06750.
 XX
 XX 30-SEP-1999; 99JP-0280098.
 PR 28-JUN-2000; 2000JP-0194043.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;
 DR NPI; 2001-266172/27.
 DR N-PSDB; AAF86283.
 XX
 XX Efficient secretory production of foreign proteins e.g.
 PT transglutaminase employing transformant coryneform bacterium, simply on
 PT industrial scale with direct recovery for use in food processing and
 PT pharmaceutical industry -
 XX
 XX Example 1; Page 90-93; 151pp; Japanese.
 XX
 XX This invention relates to a process for the production of a foreign
 CC secretory protein through the construction of a recombinant coryneform
 CC bacterium. The coryneform bacterium is transformed with an expression
 CC construct in which DNA encoding a target foreign protein pro-structure is
 CC ligated to the downstream region of DNA encoding the signal peptide
 CC domain of a coryneform bacterial protein. Following transformation with
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from
 CC the expressed protein. Transglutaminases produced using this process are
 CC useful in the food processing and pharmaceutical industries. The present
 CC sequence prepro-transglutaminase. The protein is used in an example
 CC illustrating the method of the invention.
 CC
 SQ Sequence 407 AA;
 Query Match 100.0%; Score 1811; DB 22; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRGRKQOMTEOREML 60
 DB 77 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRGRKQOMTEOREML 136
 QY 61 SYGCVGTWVNSGOYPTNRLAFASPEDERFKNELKNGRPSGSTRAPFGYAKESPDEE 120
 DB 137 SYGCVGTWVNSGOYPTNRLAFASPEDERFKNELKNGRPSGSTRAPFGYAKESPDEE 196
 QY 121 KGFQARREVASVMRRALENADHESAYLDNLKKELANGNDALRNEDARSFPYSALNTPSF 180
 DB 197 KGFQARREVASVMRRALENADHESAYLDNLKKELANGNDALRNEDARSFPYSALNTPSF 256
 QY 181 KERNGNNDPSRMKAVIYSKHFWNSQDSSSADKRYGDDPARPAPGTLVDSRDN 240
 DB 257 KERNGNNDPSRMKAVIYSKHFWNSQDSSSADKRYGDDPARPAPGTLVDSRDN 316
 QY 241 PRSPTPSGEGFVNDYGMFGAQTEDADAKTWTGHNHYHA PNGSLGAMHVYSKFRWSE 300
 DB 317 PRSPTPSGEGFVNDYGMFGAQTEDADAKTWTGHNHYHA PNGSLGAMHVYSKFRWSE 376
 QY 301 GYSDFDRGAYVITFIPKSNWNTAPDKYKQGW 331
 DB 377 GYSDFDRGAYVITFIPKSNWNTAPDKYKQGW 407

XX
 XX Coryneform bacterium; signal peptide domain; food processing; medicine;
 XX cosmetic; transglutaminase; human epithelial growth factor.
 XX
 OS Streptococcus thermophilus mobaraense.
 XX
 XX WO200261694-A1.
 XX
 PD 17-OCT-2002.
 XX
 XX 27-MAR-2002; 2002WO-JP02978.
 PR 30-MAR-2001; 2001JP-0098808.
 XX
 XX (AJIN) AJINOMOTO CO INC.
 PA
 XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Helma H, Matsui H;
 PI
 XX WPI; 2003-058550/05.
 DR N-PSDB; ABX10285.
 XX
 XX Secretion production of foreign proteins by culturing transformant
 PT coryneform bacteria, applicable in producing e.g. industrially-useful
 PT transglutaminase and human epithelial growth factor -
 XX
 XX Disclosure; Page 68-71; 117pp; Japanese.
 XX
 XX The invention relates to a method for producing a foreign protein by
 CC culturing a mutant strain of a coryneform bacterium that contains an
 CC expression gene construct to secrete the foreign protein, followed by
 CC recovery of the produced foreign protein. The gene construct is obtained
 CC by ligation of a nucleic acid encoding a signal peptide domain
 CC originating from a coryneform bacterium, to downstream of a functioning
 CC promoter sequence in the coryneform bacterium and also by ligation of a
 CC nucleic acid encoding a foreign protein, to downstream of a nucleic acid
 CC sequence encoding the signal peptide. The method is useful for the
 CC production of a foreign protein which is applicable in producing e.g.
 CC industrially-useful transglutaminase and human epithelial growth factor
 CC for use in medicine, cosmetics and food processing. This sequence
 CC represents a foreign protein of the invention.
 CC
 SQ Sequence 407 AA;
 Query Match 100.0%; Score 1811; DB 24; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRGRKQOMTEOREML 60
 DB 77 DSDDRVTTPPAEPLDRMPDPRPSYGRATVNNYIRKQOVYSHRGRKQOMTEOREML 136
 QY 61 SYGCVGTWVNSGOYPTNRLAFASPEDERFKNELKNGRPSGSTRAPFGYAKESPDEE 120
 DB 137 SYGCVGTWVNSGOYPTNRLAFASPEDERFKNELKNGRPSGSTRAPFGYAKESPDEE 196
 QY 181 KERNGNNDPSRMKAVIYSKHFWNSQDSSSADKRYGDDPARPAPGTLVDSRDN 240
 DB 257 KERNGNNDPSRMKAVIYSKHFWNSQDSSSADKRYGDDPARPAPGTLVDSRDN 316
 QY 241 PRSPTPSGEGFVNDYGMFGAQTEDADAKTWTGHNHYHA PNGSLGAMHVYSKFRWSE 300
 DB 317 PRSPTPSGEGFVNDYGMFGAQTEDADAKTWTGHNHYHA PNGSLGAMHVYSKFRWSE 376
 QY 301 GYSDFDRGAYVITFIPKSNWNTAPDKYKQGW 331
 DB 377 GYSDFDRGAYVITFIPKSNWNTAPDKYKQGW 407

AAR49049
 ID AAR49049 standard; Protein: 408 AA.
 XX AAR49049;
 AC AAR49049;
 XX
 DT 10-MAR-2003 (updated)
 DT 20-SEP-1994 (first entry)
 XX
 DE BTG-contg. sequence.
 XX
 KW Bacterial transglutaminase; BTG; expression;
 KW active; inactive; inclusion body.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN JP06030771-A.
 PD 08-FEB-1994.
 PD 14-JUL-1992; 92JP-0187038.
 PF 14-JUL-1992; 92JP-0187038.
 PR 14-JUL-1992; 92JP-0187038.
 XX
 PA (AJIN) AJINOMOTO KK.
 PA
 XX
 DR MPI; 1994-07929%/10.
 DR N-PSDB; AAQ55984.
 XX
 PT Prep'n. of bacterial trans-glutaminase in large amts. - by
 PT expression of fusion protein in E. coli bacterial
 PT trans-glutaminase
 XX
 PS Disclosure; Page 8-10; 13pp; Japanese.
 XX
 CC A fused protein contains amino acids 16-346 of BTG (AAQ55983)
 CC and a hydrophilic peptide at the amino terminal.
 CC Expression of DNA encoding this protein in E. coli allow large
 CC scale prodn. of BTG. An active BTG can be prep'd. from the
 CC inactive fused protein inclusion body.
 CC (updated on 10-MAR-2003 to add missing OS field.)
 CC
 XX
 SO Sequence 408 AA;
 Query Match 100.0%; Score 1811; DB 15; Length 408;
 Best Local Similarity 100.0%; Pred. No. 3,9e-153;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSDDRVTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHRDRKQMTBEQREVL 60
 DB 78 DSDDRVTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHRDRKQMTBEQREVL 137
 QY 61 SYGCVGVTVNNSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETRAEFEGRVAKESFDEE 120
 DB 138 SYGCVGVTVNNSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETRAEFEGRVAKESFDEE 197
 QY 121 KGFQRAAEVAVNMRALENAHDESAYLDNLKKELANGNDALINEDARSPYSALRNTPEF 180
 DB 198 KGFQRAAEVAVNMRALENAHDESAYLDNLKKELANGNDALINEDARSPYSALRNTPEF 257
 QY 161 KERNGNHDSRMKAVIYSKHFWSGDRSSADKRYGDPDAFRAPGTGLVMSDRNT 240
 DB 258 KERNGNHDSRMKAVIYSKHFWSGDRSSADKRYGDPDAFRAPGTGLVMSDRNT 317
 QY 241 PRSPTSPGEGFVNFIDYMGFGAQTEADKTVTHGNHYAPGSLGAMHYVESKFRANMSE 300
 DB 318 PRSPTSPGEGFVNFIDYMGFGAQTEADKTVTHGNHYAPGSLGAMHYVESKFRANMSE 377
 QY 301 GYSPFDGAYVITFIKSNTPADKVKQGNP 331
 DB 378 GYSPFDGAYVITFIKSNTPADKVKQGNP 408

RESULT 15
 AAR22653
 ID AAR22653 standard; Protein: 331 AA.
 XX AAR22653;
 AC AAR22653;
 XX
 DT 25-MAR-2003 (updated)
 DT 09-OCT-1992 (first entry)
 XX
 DE Transglutaminase (expressed in E. coli).
 XX
 KW BTG; acyl rearrangement; deamination; yeast; actinomycetes.
 KW
 XX
 OS Synthetic.
 OS
 PN EP481504-A.
 PD 22-APR-1992.
 PD 18-OCT-1991; 91EP-0117813.
 PF 18-OCT-1991; 91EP-0117813.
 PR 19-OCT-1990; 90JP-0282566.
 XX
 PA (AMANO) AMANO PHARM KK.
 PA (AJIN) AJINOMOTO KK.
 PA
 XX
 PI Ando K, Arakawa S, Koike S, Matsui H, Takagi H, Washizu K;
 DR MPI; 1992-133808/17.
 DR N-PSDB; AAQ24207.
 PT DNA fragment encoding trans-glutaminase - is inserted into
 PT vector, e.g. Pvu1053-BTG, for protein expression
 XX
 PS Disclosure; Page 3; 55pp; English.
 XX
 CC The protein sequence given has transglutaminase (BTG) activity. The
 CC DNA encoding this protein has a base sequence which can be used.
 CC suitably in an expression system using E. coli or yeast as a host.
 CC The base sequence can be compared to those given in AAQ24197 and
 CC AAQ24200 which are derived from Actinomyces sp.
 CC BTG catalyses an acyl rearrangement reaction of a gamma-carboxamide
 CC gp. of glutamine. It introduces intra- or intermolecular formation
 CC of epsilon-(gamma-Gln)-lys cross-linking when an epsilon-amino gp.
 CC of a lys residue acts as an acyl receptor. When water acts as an
 CC acyl acceptor the enzyme accelerates the conversion of Gln residues
 CC to Glu residues by deamination.
 CC The enzyme is used in the prodn. of gelled foods, gelled cosmetics,
 CC yogurt, gelatin, cheese etc. It is also used in the prodn. of
 CC thermally stable materials such as microcapsules and carriers of
 CC immobilized enzymes. The DNA sequence given allows the prodn. of
 CC BTG efficiently and in large quantity.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX
 SO Sequence 331 AA;
 Query Match 99.5%; Score 1802; DB 13; Length 331;
 Best Local Similarity 99.7%; Pred. No. 1.8e-152;
 Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 DSDDRTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHRDRKQMTBEQREVL 60
 DB 1 DSDDRTPPAEPLDRPDPYRPSYGRATVNNYIRKQOYVSHRDRKQMTBEQREVL 60
 QY 61 SYGCVGVTVNNSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETRAEFEGRVAKESFDEE 120
 DB 61 SYGCVGVTVNNSGOYPTNRLAFASFEDEDFKNEIKNGRPRSGETRAEFEGRVAKESFDEE 120
 QY 121 KGFQRAAEVAVNMRALENAHDESAYLDNLKKELANGNDALINEDARSPYSALRNTPEF 180
 DB 121 KGFQRAAEVAVNMRALENAHDESAYLDNLKKELANGNDALINEDARSPYSALRNTPEF 180
 QY 181 KERNGNHDSRMKAVIYSKHFWSGDRSSADKRYGDPDAFRAPGTGLVMSDRNT 240

Db 181 KERNQGNHDPSSRMKAVIYSKHFWSSQDRSSSADKKRYGDDAFRPAFGTGLVDMGRVNI 240
QY 241 PRSPTSPGEGFVNFDYGMFGAQTADADADKTWTHGNHYHAPNGSLGAMHVBKFRNWS 300
Db 241 PRSPTSPGEGFVNFDYGMFGAQTADADADKTWTHGNHYHAPNGSLGAMHVBKFRNWS 300
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331
Db 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331

Search completed: September 22, 2003, 12:43:55
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:43:10 / Search time 20 Seconds
(without alignments)
700.245 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDRYTPPAPBLDRMPDY.....ITFIPKSMNTAPDKVQGW 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA.*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5S-COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6S-COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6CTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	331	1	Sequence 1, Appl1
2	1811	100.0	331	3	Sequence 1, Appl1
3	1811	100.0	331	3	Sequence 3, Appl1
4	1811	100.0	331	3	Sequence 3, Appl1
5	1811	100.0	331	4	Sequence 1, Appl1
6	1811	100.0	406	1	Sequence 13, Appl1
7	1465	80.9	331	3	Sequence 2, Appl1
8	1465	80.9	331	3	Sequence 2, Appl1
9	591	32.6	126	3	Sequence 7, Appl1
10	591	32.6	126	3	Sequence 7, Appl1
11	208	11.5	40	3	Sequence 8, Appl1
12	208	11.5	40	3	Sequence 8, Appl1
13	189	10.4	52	3	Sequence 6, Appl1
14	189	10.4	52	3	Sequence 6, Appl1
15	102	5.6	735	4	Sequence 1, Appl1
16	99	5.5	315	4	Sequence 5, Appl1
17	93.5	5.2	1274	4	Sequence 20366, A
18	93	5.1	734	4	Sequence 4, Appl1
19	91.5	5.1	333	4	Sequence 28443, A
20	91	5.0	544	4	Sequence 7, Appl1
21	91	5.0	544	4	Sequence 15, Appl1
22	91	5.0	844	4	Sequence 25037, A
23	90.5	5.0	367	4	Sequence 30, Appl1
24	90.5	5.0	367	4	Sequence 30, Appl1
25	90.5	5.0	537	4	Sequence 27024, A
26	89.5	4.9	935	4	Sequence 25, Appl1
27	89	4.9	735	4	Sequence 18657, A

28	88	4.9	10182	4	US-09-134-001C-31E9	Sequence 3159, Ap
29	87.5	4.8	1128	4	US-08-111-939-2	Sequence 2, Appl1
30	87.5	4.8	1128	4	US-09-641-741-30	Sequence 30, Appl1
31	87.5	4.8	1128	4	US-09-060-482-8	Sequence 8, Appl1
32	86.5	4.8	913	2	US-08-487-890A-5	Sequence 5, Appl1
33	86.5	4.8	913	2	US-08-478-435-5	Sequence 5, Appl1
34	86.5	4.8	913	2	US-08-337-483-5	Sequence 5, Appl1
35	86.5	4.8	913	2	US-08-478-373-5	Sequence 5, Appl1
36	86.5	4.8	913	3	US-08-474-671-5	Sequence 5, Appl1
37	86.5	4.8	913	3	US-08-483-577A-5	Sequence 5, Appl1
38	86.5	4.8	913	3	US-08-897-438-5	Sequence 5, Appl1
39	86.5	4.8	913	4	US-08-637-654-5	Sequence 5, Appl1
40	86.5	4.8	913	4	US-08-649-518-5	Sequence 5, Appl1
41	85	4.7	778	4	US-09-198-452A-508	Sequence 508, Ap
42	84.5	4.7	427	4	US-09-134-001C-5143	Sequence 5143, Ap
43	84.5	4.7	625	3	US-09-360-197-15	Sequence 15, Appl1
44	84.5	4.7	674	3	US-08-893-852A-1	Sequence 1, Appl1
45	84.5	4.7	702	4	US-09-107-532A-6866	Sequence 6866, Ap

ALIGNMENTS

RESULT: 1
US-08-136-993-1
Sequence 1, Application US/08136993
Patent No. 542025
GENERAL INFORMATION:
APPLICANT: Takagi, Hiroshi
APPLICANT: Araioka, Shiro
APPLICANT: Matsui, Hiroshi
APPLICANT: Washizu, Kinya
APPLICANT: Ando, Keiichi
APPLICANT: Koike, Satoshi
TITLE OF INVENTION: Recombinant transglutaminase
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue
CITY: N.W.
STATE: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/136,993
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/777,447
FILING DATE:
APPLICATION NUMBER: JP 2-282566
FILING DATE: 19-OCT-1990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-293-7060
TELEFAX: 202-293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-136-993-1
Query Match 100.0%; Score 1811; DB 1; Length 331;
Best Local Similarity 100.0%; Pred. No. 9.2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEFLRMDPDPYRPSYGAETVANNYIRKQOYVSHRDGRKQOMTEBQREML 60
DB 1 DSDDRVTPPAEFLRMDPDPYRPSYGAETVANNYIRKQOYVSHRDGRKQOMTEBQREML 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDRFKNELKNGRPRSGETRAEFERVAKESFDEE 120
DB 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDRFKNELKNGRPRSGETRAEFERVAKESFDEE 120
QY 121 KGFQARREVASVNNALENAHDESAVLDNLKKELANGDALRNEDASPPYSALRNTPSF 180
DB 121 KGFQARREVASVNNALENAHDESAVLDNLKKELANGDALRNEDASPPYSALRNTPSF 180
QY 181 KERNGNNDPSSMKAVIYSKHFWSGODRSSADKKRYGDPAPRPAAGTGLVDMGRDNI 240
DB 181 KERNGNNDPSSMKAVIYSKHFWSGODRSSADKKRYGDPAPRPAAGTGLVDMGRDNI 240
QY 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAHNSGLGAMHYESKFRNMS 300
DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAHNSGLGAMHYESKFRNMS 300
QY 301 GYSDPDRGAVYITFIPKSNWITAPDKYKQGP 331
DB 301 GYSDPDRGAVYITFIPKSNWITAPDKYKQGP 331

RESULT 2

US-09-109-063-1
Sequence 1, Application US/09109063
Patent No. 6013498
GENERAL INFORMATION:
APPLICANT: YOKOYAMA, KEIICHI
APPLICANT: NAKAMURA, NAEMI
APPLICANT: MIWA, TETSUYA
APPLICANT: SEGURO, KATSUYA
TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
FILE REFERENCE: 0010-0937-0
CURRENT APPLICATION NUMBER: US/09/109,063
CURRENT FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: JP 180010/1997
EARLIER FILING DATE: 1997-07-04
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial
OTHER INFORMATION: Sequence: TRANSGLUTAMINASE
US-09-109-063-1

Query Match 100.0%; Score 1811; DB 3; Length 331;
Best Local Similarity 100.0%; Pred. No. 9.2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEFLRMDPDPYRPSYGAETVANNYIRKQOYVSHRDGRKQOMTEBQREML 60
DB 1 DSDDRVTPPAEFLRMDPDPYRPSYGAETVANNYIRKQOYVSHRDGRKQOMTEBQREML 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDRFKNELKNGRPRSGETRAEFERVAKESFDEE 120
DB 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDRFKNELKNGRPRSGETRAEFERVAKESFDEE 120
QY 121 KGFQARREVASVNNALENAHDESAVLDNLKKELANGDALRNEDASPPYSALRNTPSF 180
DB 121 KGFQARREVASVNNALENAHDESAVLDNLKKELANGDALRNEDASPPYSALRNTPSF 180
QY 181 KERNGNNDPSSMKAVIYSKHFWSGODRSSADKKRYGDPAPRPAAGTGLVDMGRDNI 240
DB 181 KERNGNNDPSSMKAVIYSKHFWSGODRSSADKKRYGDPAPRPAAGTGLVDMGRDNI 240
QY 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAHNSGLGAMHYESKFRNMS 300
DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAHNSGLGAMHYESKFRNMS 300

DB 241 PRSPTSPGEGFVNFYGFAGQTEADADKTWTHGNHYHAHNSGLGAMHYESKFRNMS 300
QY 301 GYSDPDRGAVYITFIPKSNWITAPDKYKQGP 331
DB 301 GYSDPDRGAVYITFIPKSNWITAPDKYKQGP 331

RESULT 3

US-08-793-426A-3
Sequence 3, Application US/08793426A
Patent No. 6100053
GENERAL INFORMATION:
APPLICANT: Becht, Lisbeth
APPLICANT: No. 6100053revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schafer, Thomas
TITLE OF INVENTION: Microbial Transglutaminases, Their
TITLE OF INVENTION: Production And Use
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,426A
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-793-426A-3

Query Match 100.0%; Score 1811; DB 3; Length 331;
Best Local Similarity 100.0%; Pred. No. 9.2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEFLRMDPDPYRPSYGAETVANNYIRKQOYVSHRDGRKQOMTEBQREML 60
DB 1 DSDDRVTPPAEFLRMDPDPYRPSYGAETVANNYIRKQOYVSHRDGRKQOMTEBQREML 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDRFKNELKNGRPRSGETRAEFERVAKESFDEE 120
DB 61 SYGCVGVTVWNSGQYPTNRLAFASFDDEDRFKNELKNGRPRSGETRAEFERVAKESFDEE 120
QY 121 KGFQARREVASVNNALENAHDESAVLDNLKKELANGDALRNEDASPPYSALRNTPSF 180
DB 121 KGFQARREVASVNNALENAHDESAVLDNLKKELANGDALRNEDASPPYSALRNTPSF 180
QY 181 KERNGNNDPSSMKAVIYSKHFWSGODRSSADKKRYGDPAPRPAAGTGLVDMGRDNI 240
DB 181 KERNGNNDPSSMKAVIYSKHFWSGODRSSADKKRYGDPAPRPAAGTGLVDMGRDNI 240

QY 241 PRSPSPGEGFVNFVDFGWFAGQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
DB 241 PRSPSPGEGFVNFVDFGWFAGQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
QY 301 GYSDPDRGAYVITFIPKSWNTAPDKVKQGW 331
DB 301 GYSDPDRGAYVITFIPKSWNTAPDKVKQGW 331

RESULT 4

US-09-294-565-3

; Sequence 3, Application US/09294565

; Patent No. 6190879

; GENERAL INFORMATION:

; APPLICANT: Sech, Lisbeth

; APPLICANT: No. 6190879, Iben

; APPLICANT: Halkier, Torden

; APPLICANT: Rasmussen, Grethe

; APPLICANT: Schaefer, Thomas

; APPLICANT: Andersen, Jens

; TITLE OF INVENTION: Microbial Transglutaminases, Their

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS: 10

; ADDRESS: No. 61908790 No. 61908790disk of No. 61908790h America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSRO for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/294,565

; FILING DATE: 19-APR-1999

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Green, Reza

; REGISTRATION NUMBER: 38,475

; REFERENCE/DOCKET NUMBER: 4211,224-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 331 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-294-565-3

Query Match

Best Local Similarity 100.0%; Score 1811; DB 3; Length 331;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQOYVSHDGRKQOQTEOQEM 60
DB 1 DSDDVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQOYVSHDGRKQOQTEOQEM 60
QY 61 SYGCVGVTVNSGOYPTNRLAFASPEDDFPNELKNGRPSGRTAFEGRVAKESFDEE 120
DB 61 SYGCVGVTVNSGOYPTNRLAFASPEDDFPNELKNGRPSGRTAFEGRVAKESFDEE 120
QY 121 KGFQRAREVASVNMRALENHADSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180
DB 121 KGFQRAREVASVNMRALENHADSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180

QY 181 KERNGNHDPBRMKAVYSAHFWSGODRSSADKRYGDPDAPRPAFGVLVDMSPDNI 240
DB 181 KERNGNHDPBRMKAVYSAHFWSGODRSSADKRYGDPDAPRPAFGVLVDMSPDNI 240
QY 241 PRSPSPGEGFVNFVDFGWFAGQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
DB 241 PRSPSPGEGFVNFVDFGWFAGQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
QY 301 GYSDPDRGAYVITFIPKSWNTAPDKVKQGW 331
DB 301 GYSDPDRGAYVITFIPKSWNTAPDKVKQGW 331

RESULT 5

US-09-448-310-1

; Sequence 1, Application US/09448310

; Patent No. 6538122

; GENERAL INFORMATION:

; APPLICANT: YOKOMURA, KEIICHI

; APPLICANT: NAKAMURA, NAMI

; APPLICANT: MINA, TETSUYA

; APPLICANT: SEGRO, KATSUYA

; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE

; FILE REFERENCE: 0010-0937-0

; CURRENT APPLICATION NUMBER: US/09/448,310

; PRIOR FILING DATE: 1999-11-24

; PRIOR APPLICATION NUMBER: 09/109,063

; PRIOR FILING DATE: 1998-07-02

; NUMBER OF SEQ ID NOS: 62

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 331

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial

; OTHER INFORMATION: Sequence:TRANSGLUTAMINASE

US-09-448-310-1

Query Match 100.0%; Score 1811; DB 4; Length 331;
Best Local Similarity 100.0%; Pred. No. 9,2e-171;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQOYVSHDGRKQOQTEOQEM 60
DB 1 DSDDVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQOYVSHDGRKQOQTEOQEM 60
QY 61 SYGCVGVTVNSGOYPTNRLAFASPEDDFPNELKNGRPSGRTAFEGRVAKESFDEE 120
DB 61 SYGCVGVTVNSGOYPTNRLAFASPEDDFPNELKNGRPSGRTAFEGRVAKESFDEE 120
QY 121 KGFQRAREVASVNMRALENHADSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180
DB 121 KGFQRAREVASVNMRALENHADSAYLDNLKKELANGNDALRNEDARSPFYALNTPSF 180
QY 181 KERNGNHDPBRMKAVYSAHFWSGODRSSADKRYGDPDAPRPAFGVLVDMSPDNI 240
DB 181 KERNGNHDPBRMKAVYSAHFWSGODRSSADKRYGDPDAPRPAFGVLVDMSPDNI 240
QY 241 PRSPSPGEGFVNFVDFGWFAGQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
DB 241 PRSPSPGEGFVNFVDFGWFAGQTEADADKTWTHGNHYHAPNGSLGAMHYESKFRNMSE 300
QY 301 GYSDPDRGAYVITFIPKSWNTAPDKVKQGW 331
DB 301 GYSDPDRGAYVITFIPKSWNTAPDKVKQGW 331

RESULT 6

US-08-136-993-13

; Sequence 13, Application US/08136993

; Patent No. 5420025

; GENERAL INFORMATION:

APPLICANT: Takagi, Hiroshi
APPLICANT: Araioka, Shino
APPLICANT: Matsui, Hiroshi
APPLICANT: Washizu, Kinya
APPLICANT: Ando, Keiichi
APPLICANT: Koike, Satoshi
TITLE OF INVENTION: Recombinant transglutaminase
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mason, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue
CITY: N.W.
STATE: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/136,993
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/777,447
FILING DATE:
APPLICATION NUMBER: JP 2-282566
FILING DATE: 19-OCT-1990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-293-7060
TELEFAX: 202-293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-136-993-13

Query Match 100.0%; Score 1811; DB 1; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.2e-170;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDRVTPEAPPLDMPDPYRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 60
DB 76 DSDRVTPEAPPLDMPDPYRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 135
QY 61 SYGCVGVTVWNSGQYFPNNRLAFASFEDERFKNELKNGRPPSGGTRAEFGRYAKSPDEE 120
DB 136 SYGCVGVTVWNSGQYFPNNRLAFASFEDERFKNELKNGRPPSGGTRAEFGRYAKSPDEE 195
QY 121 KQFORAREVASVNNRALENHDESAVLDNLKKELANGNDALREDEARSPFYALANTPSF 180
DB 196 KQFORAREVASVNNRALENHDESAVLDNLKKELANGNDALREDEARSPFYALANTPSF 255
QY 181 KERNGGNHDPBRMKAVIYSKHFWSGQDSSSADRKKYGDDAPRAPPGTGLVDMKDRNI 240
DB 256 KERNGGNHDPBRMKAVIYSKHFWSGQDSSSADRKKYGDDAPRAPPGTGLVDMKDRNI 315
QY 241 PRSPTSPGEGVNFVDYGMFGAQTEDADKXTVWGNHYHAPNGSLGAMHYESKFRNWS 300
DB 316 PRSPTSPGEGVNFVDYGMFGAQTEDADKXTVWGNHYHAPNGSLGAMHYESKFRNWS 375
QY 301 GYSDFDRGAYITFIIPKSNMTAPDKVKQGW 331
DB 376 GYSDFDRGAYITFIIPKSNMTAPDKVKQGW 406

RESULT 7
US-08-793-426A-2
Sequence 2, Application US/08793426A

Patent No. 6100053
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6100053revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaefer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,426A
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-793-426A-2

Query Match 80.3%; Score 1465; DB 3; Length 331;
Best Local Similarity 79.3%; Pred. No. 1.4e-136;
Matches 261; Conservative 34; Mismatches 34; Indels 0; Gaps 0;
QY 2 SDRVTPEAPPLDMPDPYRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 61
DB 2 ADRVTPEAPPLDMPDPYRPSYGRATVYNNYIRKMQQVYSHRDGRKQWTEBQREWL 61
QY 62 YGCVGVTVWNSGQYFPNNRLAFASFEDERFKNELKNGRPPSGGTRAEFGRYAKSPDEE 121
DB 62 YGCVGVTVWNSGQYFPNNRLAFASFEDERFKNELKNGRPPSGGTRAEFGRYAKSPDEE 121
QY 122 GFQARAREVASVNNRALENHDESAVLDNLKKELANGNDALREDEARSPFYALANTPSF 181
DB 122 GFQARAREVASVNNRALENHDESAVLDNLKKELANGNDALREDEARSPFYALANTPSF 181
QY 182 ERNGGNHDPBRMKAVIYSKHFWSGQDSSSADRKKYGDDAPRAPPGTGLVDMKDRNI 241
DB 182 ERNGGNHDPBRMKAVIYSKHFWSGQDSSSADRKKYGDDAPRAPPGTGLVDMKDRNI 241
QY 242 RSPSPGEGVNFVDYGMFGAQTEDADKXTVWGNHYHAPNGSLGAMHYESKFRNWS 301
DB 242 RSPSPGEGVNFVDYGMFGAQTEDADKXTVWGNHYHAPNGSLGAMHYESKFRNWS 301
QY 302 YSDFDRGAYITFIIPKSNMTAPDKVKQGW 330
DB 302 YSDFDRGAYITFIIPKSNMTAPDKVKQGW 330

RESULT 8

US-09-294-565-2

/ Sequence 2, Application US/09294565
/ Patent No. 6190879
/ GENERAL INFORMATION:
/ APPLICANT: Bech, Lisbeth
/ APPLICANT: No. 6190879evang, Iben
/ APPLICANT: Halkier, Torben
/ APPLICANT: Rasmussen, Grethe
/ APPLICANT: Schafer, Thomas
/ APPLICANT: Andersen, Jens
/ TITLE OF INVENTION: Microbial Transglutaminases, Their
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: No. 6190879c No. 6190879disk of No. 6190879ch America, Inc.
/ STREET: 405 Lexington Avenue
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10174
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/294,565
/ FILING DATE: 19-APR-1999
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Green, Reza
/ REGISTRATION NUMBER: 38,475
/ REFERENCE/DOCKET NUMBER: 4211.224-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-867-0123
/ TELEFAX: 212-878-9655
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 331 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-294-565-2

Query Match 80.9%; Score 1465; DB 3; Length 331;
Best Local Similarity 79.3%; Pred. No. 1,4e-136;
Matches 261; Conservative 34; Mismatches 34; Indels 0; Gaps 0;

QY 2 SDRVTTPPAEPLDNDPPTPSYGAETVNNYIRKQOQVSHRDRKQOMTEOREKLS 61
DB 2 ADRVTTPPAEPLDNDPPTPSYGAETVNNYIRKQOQVSHRDRKQOMTEOREKLS 61
QY 62 YGVGVTVNWSGQVTVNRLAFASFDPRFKNLKXGRPSGERTRAEFERVAKSFDEEK 121
DB 62 YGVGVTVNWSGQVTVNRLAFASFDPRFKNLKXGRPSGERTRAEFERVAKSFDEEK 121
QY 122 GFORAEVAVSWMRALNADHESAYLDNKKELANGNDALNEDASPPYSALRNPESFK 181
DB 122 GFORAEVAVSWMRALNADHESAYLDNKKELANGNDALNEDASPPYSALRNPESFK 181
QY 182 ERGAGNDPSRKATVSKHFWSGDRSSSAKRYGPDAPFRACGTGLVMSRDRIIP 241
DB 182 ERGAGNDPSRKATVSKHFWSGDRSSSAKRYGPDAPFRACGTGLVMSRDRIIP 241
QY 242 RSPDTSRGEFVFDYGVFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFPNNSEG 301
DB 242 RSPDTSRGEFVFDYGVFGAQTADADKTWTHGNHYAPNGSLGAMHYESKFPNNSEG 301
QY 302 YSDFDGAIVTTFIKSWNTAPDKVYKQW 330
DB 302 YADFDGTVITFIPIKSWNTAPAEVYKQW 330

RESULT 9

US-08-793-426A-7
/ Sequence 7, Application US/08793426A
/ Patent No. 6100053
/ GENERAL INFORMATION:
/ APPLICANT: Bech, Lisbeth
/ APPLICANT: No. 6100053evang, Iben
/ APPLICANT: Halkier, Torben
/ APPLICANT: Rasmussen, Grethe
/ APPLICANT: Schafer, Thomas
/ APPLICANT: Andersen, Jens
/ TITLE OF INVENTION: Microbial Transglutaminases, Their
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: No. 6100053c No. 6100053disk of No. 6100053ch America, Inc.
/ STREET: 405 Lexington Avenue
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10174
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/793,426A
/ FILING DATE: 25-FEB-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rozek, Carol B.
/ REGISTRATION NUMBER: 36,993
/ REFERENCE/DOCKET NUMBER: 4211.204-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-867-0123
/ TELEFAX: 212-878-9655
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 126 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/ US-08-793-426A-7

Query Match 32.6%; Score 591; DB 3; Length 126;
Best Local Similarity 82.5%; Pred. No. 8e-51;
Matches 104; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 165 DAREPFGALRNPSPSKERNGANDPSRKAVVYSKHFSGODRSSADRRKGDPAAR 224
DB 1 DRRSSFGALRNPSPSKERNGANDPSRKAVVYSKHFSGODRSSADRRKGDPAAR 60
QY 225 PAPDTGLVDSRDRIIPSPSPSPSGEVNFDYGVFGAQTADADKTWTHGNHYAPNGS 284
DB 61 PDGDTGLVDSRDRIIPSPSPSPSGEVNFDYGVFGAQTADADKTWTHGNHYAPNGS 120
QY 285 LGANHV 290
DB 121 MGPMBV 126

RESULT 10

US-09-294-565-7
/ Sequence 7, Application US/09294565
/ Patent No. 6190879
/ GENERAL INFORMATION:

TITLE OF INVENTION: Microbial Transglutaminases, Their
 TITLE OF INVENTION: Production And Use
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053ch America, Inc.
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/793,426A
 FILING DATE: 25-FEB-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rozek, Carol E.
 REGISTRATION NUMBER: 36,993
 REFERENCE/DOCKET NUMBER: 421.1.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 40 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: C-terminal
 US-08-793-426A-8

Query Match 11.5%; Score 208; DB 3; Length 40;
 Best Local Similarity 90.0%; Pred. No. 1e-13;
 Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0

QY 292 ESKFRNWSGYSDFDKGAVYTFIFPKSWNTAPDKVQGNP 331
 |||||
 DB 1 ESKFRNWSAGYADPDKGAVYTFIFPKSWNTAPARVQGNP 40
 |||||

RESULT 12
 US-09-294-565-8
 Sequence 8, Application US/09294565
 Patent No. 6190879
 GENERAL INFORMATION:
 APPLICANT: Bech, Lisbeth
 APPLICANT: No. 6190879revang, Iben
 APPLICANT: Halkier, Torben
 APPLICANT: Rasmussen, Gertne
 APPLICANT: Schaefer, Thomas
 APPLICANT: Andersen, Jens
 TITLE OF INVENTION: Microbial Transglutaminases, Their
 TITLE OF INVENTION: Production And Use
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: No. 6190879 No. 6190879disk of No. 6190879ch America, Inc.
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS


```
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/294,565
FILING DATE: 19-APR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4211.224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-09-294-565-8

Query Match          11.5%; Score 208; DB 3; Length 40;
Best Local Similarity 90.0%; Pred. No. 1e-13;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      292 ESKFRNWSBGYSDFDRAGAYITIPKSNMTAPDKYKQGP 331
Db      1 ESKFRNWSAGYADPDGAYITIPKSNMTAPAEVKQGP 40

RESULT 13
US-08-793-426A-6
Sequence 6, Application US/08793426A
Patent No. 6100053
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6100053revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaefer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 61000530 No. 6100053disk of No. 6100053ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,426A
FILING DATE: 25-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 6:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 52 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
US-08-793-426A-6

Query Match          10.4%; Score 189; DB 3; Length 52;
Best Local Similarity 70.6%; Pred. No. 1.1e-11;
Matches 36; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY      106 AEFGRVAKESFDEKGFQARASVAVNRALENHDSAYLDNLKTELAN 156
Db      1 AEFGRVAKGFADRGKRGARASVAVNRALENHDSAYLDNLKTELAN 51

RESULT 14
US-09-294-565-6
Sequence 6, Application US/09294565
Patent No. 6190879
GENERAL INFORMATION:
APPLICANT: Bech, Lisbeth
APPLICANT: No. 6190879revang, Iben
APPLICANT: Halkier, Torben
APPLICANT: Rasmussen, Grethe
APPLICANT: Schaefer, Thomas
APPLICANT: Andersen, Jens
TITLE OF INVENTION: Microbial Transglutaminases, Their
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 61908790 No. 6190879disk of No. 6190879ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/294,565
FILING DATE: 19-APR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4211.224-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
US-09-294-565-6

Query Match          10.4%; Score 189; DB 3; Length 52;
Best Local Similarity 70.6%; Pred. No. 1.1e-11;
Matches 36; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY      106 AEFGRVAKESFDEKGFQARASVAVNRALENHDSAYLDNLKTELAN 156
```

Db 1 AEFGRKAGKGFDAFGRGFAREVSVNNKALDSADHDEGTVIDHKLTELAV 51

RESULT 15
US-09-321-589-1
; Sequence 1, Application US/09321589
; Patent No. 6498244
; GENERAL INFORMATION:
; APPLICANT: PATEL, SALIL D.
; APPLICANT: MCARTHUR, JAMES G.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
; FILE REFERENCE: 39672
; CURRENT APPLICATION NUMBER: US/09/321,589
; CURRENT FILING DATE: 1999-05-28
; NUMBER OF SEQ. ID NOS.: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ. ID NO. 1
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Adeno-associated virus
US-09-321-589-1

Query Match 5.6%; Score 102; DB 4; Length 735;
Best Local Similarity 22.7%; Pred. No. 0.24;
Matches 77; Conservative 44; Mismatches 154; Indels 64; Gaps 22;

QY 18 DPYPSYGRAETVNNYIRKQOVSHRDGRKQOMTEOREWLSYGCYGVTVVNSGQY-- 75
Db 4 DGYLPDW--LEDTLSEGTFRQWKU---KPGPPPKPAERHKDSDRGVLPGYKYLGPFG 58
QY 76 ----PTNRLAFASFDEDR-EKNEIKNG-RP--RSGETRAEFEGRAVESFDEEKGFQRA 126
Db 59 LDKGEPVNEADAAALEHDYKAYDROLDSQDNPYLKYNHADAEFOERL-KE--DTSPGNTLG 115
QY 127 REVASVNNRRLAHNAHDESAYLDNLKKELANGNDALNEDARSFPYSALRNPSTFKERNG- 185
Db 116 RAVQAKKRVLEPL-----GLVEEPVKTPAGKKRPVEHSFV-----EPDSSSGTGK 161
QY 186 GNHDPSEMKAVIYSKHFWSGQDRSSSADKRYGDPDAFRPAD-GTGLVMSRDRNIPRSP 244
Db 162 AGQGPARKRL-----NPGOTGADSVDPDQPGPRA--APSGLGTVMTATSGAEMAD 213
QY 245 TSPG-----EGFVNFDYGVFGAQTEDADKTVW---THGNH-YHAPNGSLGAMHYAS 293
Db 214 INEGADGVNNSGNMWCDSITWMDRVITTSRT-WALPTYNHLYKOISOSGASN--DN 270
QY 294 KFRWSE--GYSDPDRGAYVITFIPKSNVNTAPDKVQGW 330
Db 271 HYFGYSTPQGIYDFDNR--FHCHFSRDMQRL--INNWN 304

Search completed: September 22, 2003, 12:45:54
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:45:10 ; Search time 62 Seconds

(without alignments)
794.898 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDRVTPEPAPLDRMPDY.....ITFPMKMTAPDKVKGWP 331

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/prodata/2/pubpa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubpa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubpa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubpa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/2/pubpa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubpa/US07_PUBCOMB.pep:*
7: /cgn2_6/prodata/2/pubpa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubpa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubpa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubpa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubpa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubpa/US09C_PUBCOMB.pep:*
13: /cgn2_6/prodata/2/pubpa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubpa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubpa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubpa/US10_NEW_PUB.pep:*
17: /cgn2_6/prodata/2/pubpa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubpa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1811	100.0	331	10	US-09-996-561-1
2	1811	100.0	331	10	US-09-884-948-1
3	1811	100.0	331	15	US-10-112-488-5
4	1811	100.0	332	9	US-09-892-864A-2
5	1811	100.0	407	14	US-10-124-429-4
6	1811	100.0	407	15	US-10-112-488-13
7	1811	95.3	410	15	US-10-022-809-2
8	1487	82.1	330	15	US-10-112-488-43
9	1487	82.1	416	14	US-10-124-429-2
10	122.5	6.8	738	12	US-10-291-583-92
11	122.5	6.8	738	12	US-10-291-583-93
12	122.5	6.8	738	12	US-10-291-583-94
13	122	6.7	737	12	US-10-291-583-95
14	117	6.5	736	12	US-10-291-583-99
15	117	6.5	737	12	US-10-291-583-72

16	114.5	6.3	1217	12	US-10-311-406-2	Sequence 2, Appl 1
17	114	6.3	735	12	US-10-291-583-68	Sequence 68, Appl 1
18	114	6.3	735	12	US-10-291-583-69	Sequence 69, Appl 1
19	113.5	6.3	729	12	US-10-291-583-111	Sequence 111, Appl 1
20	113.5	6.3	733	12	US-10-291-583-60	Sequence 60, Appl 1
21	113.5	6.3	733	12	US-10-291-583-62	Sequence 62, Appl 1
22	113.5	6.3	738	12	US-10-291-583-95	Sequence 95, Appl 1
23	112.5	6.2	733	12	US-10-291-583-61	Sequence 61, Appl 1
24	112	6.2	733	12	US-10-291-583-86	Sequence 86, Appl 1
25	112	6.2	736	12	US-10-291-583-96	Sequence 96, Appl 1
26	112	6.2	736	12	US-10-291-583-97	Sequence 97, Appl 1
27	112	6.2	736	12	US-10-291-583-98	Sequence 98, Appl 1
28	111.5	6.2	729	12	US-10-291-583-109	Sequence 109, Appl 1
29	111.5	6.2	729	12	US-10-291-583-110	Sequence 110, Appl 1
30	111.5	6.2	738	12	US-10-291-583-78	Sequence 78, Appl 1
31	111.5	6.2	738	12	US-10-291-583-80	Sequence 80, Appl 1
32	111.5	6.2	738	12	US-10-291-583-81	Sequence 81, Appl 1
33	111.5	6.2	738	12	US-10-291-583-82	Sequence 82, Appl 1
34	111	6.1	728	12	US-10-291-583-103	Sequence 103, Appl 1
35	111	6.1	728	12	US-10-291-583-104	Sequence 104, Appl 1
36	110	6.1	728	12	US-10-291-583-108	Sequence 108, Appl 1
37	110	6.1	736	12	US-10-291-583-71	Sequence 71, Appl 1
38	109.5	6.0	738	12	US-10-291-583-83	Sequence 83, Appl 1
39	109.5	6.0	738	12	US-10-291-583-84	Sequence 84, Appl 1
40	109.5	6.0	738	12	US-10-291-583-85	Sequence 85, Appl 1
41	109	6.0	728	12	US-10-291-583-101	Sequence 101, Appl 1
42	108	6.0	728	12	US-10-291-583-102	Sequence 102, Appl 1
43	108	6.0	728	12	US-10-291-583-105	Sequence 105, Appl 1
44	108	6.0	733	12	US-10-291-583-90	Sequence 90, Appl 1
45	108	6.0	1005	9	US-09-925-301-1335	Sequence 1335, Appl 1

ALIGNMENTS

RESULT 1
US-09-996-561-1
Sequence 1, Application US/0996561
Patent No. US20020151703A1
GENERAL INFORMATION:
APPLICANT: YOKOYAMA, KEIICHI
APPLICANT: YOKOYAMA, NAAMI
APPLICANT: MIYA, TETSUYA
APPLICANT: SUGURO, KATSUYA
TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
FILE REFERENCE: 0010-0937-0
CURRENT APPLICATION NUMBER: US/09/996,561
CURRENT FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/448,310
PRIOR FILING DATE: CURRENT FILING DATE: 1999-11-24
PRIOR APPLICATION NUMBER: 09/109,063
PRIOR FILING DATE: 1998-07-02
NUMBER OF SEQ ID NOS: 62
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial
OTHER INFORMATION: Sequence: TRANSGLUTAMINASE
US-09-996-561-1

Query Match 100.0%; Score 1811; DB 10; Length 331;

Best Local Similarity 100.0%; Pred. No. 5; 5e-167; Indels 0; Gaps 0;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDRVTPEPAPLDRMPDYPSYGRATVYNNYTRKQOYVSHRDRKQOMTEBQREWL 60
DB 1 DSDRVTPEPAPLDRMPDYPSYGRATVYNNYTRKQOYVSHRDRKQOMTEBQREWL 60
QY 61 SYGCVGTWNSGQYFTRRLAFASGDRKQELKNGSPRGFTFAEFERVAKESFDEE 120

```
Db 61 SYGCVGVTWVNSGQYPTNRLAFASFEDEPRFKNELKNGRPSGSETPAEFEGRVAKESFDEE 120
Qy 121 KGFQARAEVASVNNRRLAENAHDESAVLDNLKKELANGDALRNEBARSPPFSALRNTPSF 180
Db 121 KGFQARAEVASVNNRRLAENAHDESAVLDNLKKELANGDALRNEBARSPPFSALRNTPSF 180
Qy 181 KERNGNHDPSRKMAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
Db 181 KERNGNHDPSRKMAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
Qy 241 PRSPFSGEGFVNFDFYGMFGAQTBADAKTWTGHNHYHAPNGSLGAMHYESKFRWSE 300
Db 241 PRSPFSGEGFVNFDFYGMFGAQTBADAKTWTGHNHYHAPNGSLGAMHYESKFRWSE 300
Qy 301 GYSDPDRGAVITFIPKSWNTAPDKVXQGMP 331
Db 301 GYSDPDRGAVITFIPKSWNTAPDKVXQGMP 331
```

```
RESULT 2
US-09-884-948-1
; Sequence 1, Application US/09884948
; Patent No. US20020173021A1
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, KEIICHI
; APPLICANT: NAKAMURA, NAHI
; APPLICANT: MIWA, TETSUYA
; APPLICANT: SEGURO, KATSUYA
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
; FILE REFERENCE: 0010-0937-0
; CURRENT FILING DATE: 2001-06-21
; PRIOR FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: 09/448,310
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence:TRANSGLUTAMINASE
US-09-884-948-1
```

```
Query Match 100.0%; Score 1811; DB 10; Length 331;
Best Local Similarity 100.0%; Pred. No. 5,5e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 DSDRVTTPPAEPIDRMPDPYRPSYGRATVYNNYIRKMOQVYSHRDGRKQOMTEBQEWL 60
Db 1 DSDRVTTPPAEPIDRMPDPYRPSYGRATVYNNYIRKMOQVYSHRDGRKQOMTEBQEWL 60
Qy 61 SYGCVGVTWVNSGQYPTNRLAFASFEDEPRFKNELKNGRPSGSETPAEFEGRVAKESFDEE 120
Db 61 SYGCVGVTWVNSGQYPTNRLAFASFEDEPRFKNELKNGRPSGSETPAEFEGRVAKESFDEE 120
Qy 121 KGFQARAEVASVNNRRLAENAHDESAVLDNLKKELANGDALRNEBARSPPFSALRNTPSF 180
Db 121 KGFQARAEVASVNNRRLAENAHDESAVLDNLKKELANGDALRNEBARSPPFSALRNTPSF 180
Qy 181 KERNGNHDPSRKMAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
Db 181 KERNGNHDPSRKMAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
Qy 241 PRSPFSGEGFVNFDFYGMFGAQTBADAKTWTGHNHYHAPNGSLGAMHYESKFRWSE 300
Db 241 PRSPFSGEGFVNFDFYGMFGAQTBADAKTWTGHNHYHAPNGSLGAMHYESKFRWSE 300
Qy 301 GYSDPDRGAVITFIPKSWNTAPDKVXQGMP 331
Db 301 GYSDPDRGAVITFIPKSWNTAPDKVXQGMP 331
```

```
RESULT 3
US-10-112-488-5
; Sequence 5, Application US/10112488
; Publication No. US20030082746A1
; GENERAL INFORMATION:
; APPLICANT: KIKUCHI, Yoshimi
; APPLICANT: DATE, Masayo
; APPLICANT: UMEZAWA, Yukiko
; APPLICANT: YOKOYAMA, Keiichi
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
; FILE REFERENCE: 219286USOCNT
; CURRENT FILING DATE: US/10/112,488
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: PCT/JP00/06780
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: JP2000-280098
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: JP11-280098
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 331
; TYPE: PRF
; ORGANISM: Streptococcus thermophilus
US-10-112-488-5
```

```
Query Match 100.0%; Score 1811; DB 15; Length 331;
Best Local Similarity 100.0%; Pred. No. 5,5e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 DSDRVTTPPAEPIDRMPDPYRPSYGRATVYNNYIRKMOQVYSHRDGRKQOMTEBQEWL 60
Db 1 DSDRVTTPPAEPIDRMPDPYRPSYGRATVYNNYIRKMOQVYSHRDGRKQOMTEBQEWL 60
Qy 61 SYGCVGVTWVNSGQYPTNRLAFASFEDEPRFKNELKNGRPSGSETPAEFEGRVAKESFDEE 120
Db 61 SYGCVGVTWVNSGQYPTNRLAFASFEDEPRFKNELKNGRPSGSETPAEFEGRVAKESFDEE 120
Qy 121 KGFQARAEVASVNNRRLAENAHDESAVLDNLKKELANGDALRNEBARSPPFSALRNTPSF 180
Db 121 KGFQARAEVASVNNRRLAENAHDESAVLDNLKKELANGDALRNEBARSPPFSALRNTPSF 180
Qy 181 KERNGNHDPSRKMAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
Db 181 KERNGNHDPSRKMAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVMSRDRI 240
Qy 241 PRSPFSGEGFVNFDFYGMFGAQTBADAKTWTGHNHYHAPNGSLGAMHYESKFRWSE 300
Db 241 PRSPFSGEGFVNFDFYGMFGAQTBADAKTWTGHNHYHAPNGSLGAMHYESKFRWSE 300
Qy 301 GYSDPDRGAVITFIPKSWNTAPDKVXQGMP 331
Db 301 GYSDPDRGAVITFIPKSWNTAPDKVXQGMP 331
```

```
RESULT 4
US-09-892-864A-2
; Sequence 2, Application US/09892864A
; Patent No. US20020090675A1
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, Keiichi
; APPLICANT: ONO, Kunio
; APPLICANT: EJIMA, Daisuke
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
; FILE REFERENCE: 209524USOCNT
; CURRENT FILING DATE: US/09/892,864A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/JP99/07250
; PRIOR FILING DATE: 1999-12-24
; PRIOR APPLICATION NUMBER: JP 10-373131
```

PRIOR FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 332
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Synthetic Peptide
US-09-892-864A-2

Query Match 100.0%; Score 1811; DB 9; Length 332;
Best Local Similarity 100.0%; Pred. No. 5.5e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPDPYPSYGRAETVANNYIRKQOYVSHRDRKQOQWTEBQREWL 60
DB DSDDRVTPPAEPLDMPPDPYPSYGRAETVANNYIRKQOYVSHRDRKQOQWTEBQREWL 61
QY 61 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFGRVAKESFDEE 120
DB 62 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFGRVAKESFDEE 121
QY 121 KGFQAREVASYVNNALFNADHESAYLDNLKKELANGDALRNEDARSFFYALRNTPSF 180
DB 122 KGFQAREVASYVNNALFNADHESAYLDNLKKELANGDALRNEDARSFFYALRNTPSF 181
QY 181 KERNGNHDPSSRKAVIYSKHFMSGQDRSSADKRYGDPDAFPAPAGTGLVMSRDRNI 240
DB 182 KERNGNHDPSSRKAVIYSKHFMSGQDRSSADKRYGDPDAFPAPAGTGLVMSRDRNI 241
QY 241 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNWS 300
DB 242 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNWS 301
QY 301 GYSDPDRGAVITTFIPKSMNTAPDKYQGW 331
DB 302 GYSDPDRGAVITTFIPKSMNTAPDKYQGW 332

RESULT 5

US-10-124-429-4
Sequence 4, Application US/10124429
Publication No. US20020187525A1
GENERAL INFORMATION:
APPLICANT: TAGUCHI, SEIICHI
APPLICANT: MOMOSE, HARUO
TITLE OF INVENTION: A METHOD OF PRODUCING MICROBIAL TRANSGLUTAMINASE
FILE REFERENCE: 220567USOCONT
CURRENT APPLICATION NUMBER: US/10124,429
CURRENT FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: PCT/JPO0/07135
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: JP 11-295649
PRIOR FILING DATE: 1999-10-18
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 407
TYPE: PRT
ORGANISM: Streptococcus mobaraense
US-10-124-429-4

Query Match 100.0%; Score 1811; DB 14; Length 407;
Best Local Similarity 100.0%; Pred. No. 7.4e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPDPYPSYGRAETVANNYIRKQOYVSHRDRKQOQWTEBQREWL 60
DB 77 DSDDRVTPPAEPLDMPPDPYPSYGRAETVANNYIRKQOYVSHRDRKQOQWTEBQREWL 136
QY 61 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFGRVAKESFDEE 120

DB 137 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFGRVAKESFDEE 196
QY 121 KGFQAREVASYVNNALFNADHESAYLDNLKKELANGDALRNEDARSFFYALRNTPSF 180
DB 197 KGFQAREVASYVNNALFNADHESAYLDNLKKELANGDALRNEDARSFFYALRNTPSF 256
QY 181 KERNGNHDPSSRKAVIYSKHFMSGQDRSSADKRYGDPDAFPAPAGTGLVMSRDRNI 240
DB 257 KERNGNHDPSSRKAVIYSKHFMSGQDRSSADKRYGDPDAFPAPAGTGLVMSRDRNI 316
QY 241 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNWS 300
DB 317 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNWS 376
QY 301 GYSDPDRGAVITTFIPKSMNTAPDKYQGW 331
DB 377 GYSDPDRGAVITTFIPKSMNTAPDKYQGW 407

RESULT 6

US-10-112-488-13
Sequence 13, Application US/10112488
Publication No. US20030082746A1
GENERAL INFORMATION:
APPLICANT: KIKUCHI, Yoshimi
APPLICANT: DATE, Masayo
APPLICANT: UMEZAWA, Yukiko
APPLICANT: YOKOYAMA, Keiichi
APPLICANT: MATSUI, Hiroshi
TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
FILE REFERENCE: 219286USOCONT
CURRENT APPLICATION NUMBER: US/10112,488
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: PCT/JPO0/06780
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: JP2000-280098
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: JP11-280098
PRIOR FILING DATE: 1999-09-30
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 407
TYPE: PRT
ORGANISM: Streptococcus mobaraense
US-10-112-488-13

Query Match 100.0%; Score 1811; DB 15; Length 407;
Best Local Similarity 100.0%; Pred. No. 7.4e-167;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDMPPDPYPSYGRAETVANNYIRKQOYVSHRDRKQOQWTEBQREWL 60
DB 77 DSDDRVTPPAEPLDMPPDPYPSYGRAETVANNYIRKQOYVSHRDRKQOQWTEBQREWL 136
QY 61 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFGRVAKESFDEE 120
DB 137 SYGCVGTWVNSGOYPTNRLAFASFEDEFRKNELKNGRPRSGETRAEFGRVAKESFDEE 196
QY 121 KGFQAREVASYVNNALFNADHESAYLDNLKKELANGDALRNEDARSFFYALRNTPSF 180
DB 197 KGFQAREVASYVNNALFNADHESAYLDNLKKELANGDALRNEDARSFFYALRNTPSF 256
QY 181 KERNGNHDPSSRKAVIYSKHFMSGQDRSSADKRYGDPDAFPAPAGTGLVMSRDRNI 240
DB 257 KERNGNHDPSSRKAVIYSKHFMSGQDRSSADKRYGDPDAFPAPAGTGLVMSRDRNI 316
QY 241 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNWS 300
DB 317 PRSPTSGEGFVNFYGFAGQTEADADKTWTHGNHYHAENSLGAMHYESKFRNWS 376
QY 301 GYSDPDRGAVITTFIPKSMNTAPDKYQGW 331

Db 377 GYSDFDRAAYVITTFIPKSMNTAPDKYKQGW 407

RESULT 7

US-10-022-809-2

Sequence 2, Application US/10022809

Publication No. US20030113407A1

GENERAL INFORMATION:

APPLICANT: LIN, Yi-Shin

APPLICANT: LIU, Chang-Hsiueh

APPLICANT: CHU, Wen-Shen

TITLE OF INVENTION: TRANSGUTAMINASE GENE OF STREPTOCOCCUS LADAKANUM AND THE

FILE REFERENCE: U-013779-2

CURRENT APPLICATION NUMBER: US/10/022,809

CURRENT FILING DATE: 2002-03-29

PRIOR APPLICATION NUMBER: US 10/021,678

PRIOR FILING DATE: 2001-12-12

NUMBER OF SEQ ID NOS: 3

SEQ ID NO 2

LENGTH: 410

TYPE: PR

ORGANISM: Streptococcus lactis

US-10-022-809-2

Query Match 95.3%; Score 1726; DB 15; Length 410;

Best Local Similarity 93.4%; Pred. No. 1.3e-158;

Matches 309; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

Db 1 DSDRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDKQOMTEBDEE 60

Db 80 DSDRTTPAEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDKQOMTEBDEE 139

QY 61 SYGCVTVWNSGQYPTNRLAFASFDDEPRKNEKXGRPSGERTAFEBGRVAKESFDEB 120

Db 140 SYGCVTVWNSGQYPTNRLAFASFDDEPRKNEKXGRPSGERTAFEBGRVAKESFDEB 199

QY 121 KFORAREVASYVNNALNADHESAYLDNKKELANGNDALNEDASPFYSALRNTPSF 180

Db 200 KFORAREVASYVNNALNADHESAYLDNKKELANGNDALNEDASPFYSALRNTPSF 259

QY 161 KERNGNHDPSEKMAVYISKHFWGQDRSSADKRYGDPAPFAPGCTLVMSGDRNI 240

Db 260 KERNGNHDPSEKMAVYISKHFWGQDRSSADKRYGDPAPFAPGCTLVMSGDRNI 319

QY 241 PRSPITSGEGFVNFYGFAGQTEADADKTWTHGNHYHAPNSLGAMHYEESKFRNMSB 300

Db 320 PRSPITSGEGFVNFYGFAGQTEADADKTWTHGNHYHAPNSLGAMHYEESKFRNMSB 379

QY 301 GYSDFDRAAYVITTFIPKSMNTAPDKYKQGW 331

Db 380 GYSDFDRAAYVITTFIPKSMNTAPDKYKQGW 410

RESULT 8

US-10-112-488-43

Sequence 43, Application US/10112488

Publication No. US20030082746A1

GENERAL INFORMATION:

APPLICANT: KIKUCHI, Yoshimi

APPLICANT: DATE, Masayo

APPLICANT: UMEZAWA, Yukiho

APPLICANT: YOKOYAMA, Keiichi

APPLICANT: MATSUI, Hiroshi

TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGUTAMINASE

FILE REFERENCE: 219286USCONT

CURRENT APPLICATION NUMBER: US/10/112,488

CURRENT FILING DATE: 2002-04-01

PRIOR APPLICATION NUMBER: PCT/JP00/06780

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: JP2000-280038

PRIOR FILING DATE: 2000-06-28

PRIOR APPLICATION NUMBER: JP11-280098

Db 377 GYSDFDRAAYVITTFIPKSMNTAPDKYKQGW 407

RESULT 9

US-10-124-429-2

Sequence 2, Application US/10124429

Publication No. US20020187525A1

GENERAL INFORMATION:

APPLICANT: TAGUCHI, SEIICHI

APPLICANT: MOMOSE, HARUO

TITLE OF INVENTION: A METHOD OF PRODUCING MICROBIAL TRANSGUTAMINASE

FILE REFERENCE: 220567USCONT

CURRENT APPLICATION NUMBER: US/10/124,429

CURRENT FILING DATE: 2002-04-18

PRIOR APPLICATION NUMBER: PCT/JP00/07135

PRIOR FILING DATE: 2000-10-13

PRIOR APPLICATION NUMBER: JP 11-295649

PRIOR FILING DATE: 1999-10-18

NUMBER OF SEQ ID NOS: 7

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 416

TYPE: PR

ORGANISM: Streptococcus cinamomeum

US-10-124-429-2

Query Match 82.1%; Score 1487; DB 15; Length 416;

Best Local Similarity 81.5%; Pred. No. 1.8e-135;

Matches 269; Conservative 26; Mismatches 35; Indels 0; Gaps 0;

Db 2 SDRVTPPEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDKQOMTEBDEE 61

Db 87 SDRVTPPEPLDMPDPYPSYGRATYVNNYIRKQOYSHRDKQOMTEBDEE 146

QY 62 YGCVTVWNSGQYPTNRLAFASFDDEPRKNEKXGRPSGERTAFEBGRVAKESFDEB 121

Db 61 YGCVTVWNSGQYPTNRLAFASFDDEPRKNEKXGRPSGERTAFEBGRVAKESFDEB 120

QY 122 GFORAREVASYVNNALNADHESAYLDNKKELANGNDALNEDASPFYSALRNTPSF 181

Db 121 GFORAREVASYVNNALNADHESAYLDNKKELANGNDALNEDASPFYSALRNTPSF 180

QY 182 ERNGNHDPSEKMAVYISKHFWGQDRSSADKRYGDPAPFAPGCTLVMSGDRNI 241

Db 181 ERNGNHDPSEKMAVYISKHFWGQDRSSADKRYGDPAPFAPGCTLVMSGDRNI 240

QY 242 RSPITSGEGFVNFYGFAGQTEADADKTWTHGNHYHAPNSLGAMHYEESKFRNMSB 301

Db 241 RSPITSGEGFVNFYGFAGQTEADADKTWTHGNHYHAPNSLGAMHYEESKFRNMSB 300

QY 302 YSDFDRAAYVITTFIPKSMNTAPDKYKQGW 331

Db 301 YSDFDRAAYVITTFIPKSMNTAPDKYKQGW 330

Db 147 YGCVTVWNSGQYPTNRLAFASFDDEPRKNEKXGRPSGERTAFEBGRVAKESFDEB 206

```
QY 122 GPGARAVASVNNALNADHSAVLDNKKELANGDALRNEDASPTVYALNTPSFK 181
DB 207 GPGARADVASVNNALNADHSGTYINNLIKELTNNDALLRDRSNTSFK 266
QY 182 ERNGNDHPSNRKAVITSKFFMSGQDRSSADKRYGDPAPFAPAGTGLVDMGRDNIP 241
DB 267 ERNGNDHPSNRKAVITSKFFMSGQDRSSADKRYGDPAPFAPAGTGLVDMGRDNIP 326
QY 242 KSEFTSGGFFVNFYGMFGAQTADADKTWTHGNNHYAHNPSGLAMHYBSKPRNMSG 301
DB 327 RSPARSGGFFVNFYGMFGAQTADADKTWTHGNNHYAHNPSGLAMHYBSKPRNMSG 386
QY 302 YSPDRGAVITLTPKSNMTAPDKYQGM 331
DB 387 YAPDRGAVITLTPKSNMTAPDKYQGM 416

RESULT 10
US-10-291-583-92
; Sequence 92, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 92
; LENGTH: 738
; TYPE: PRT
; ORGANISM: capsid protein of AAV serotype, clone 43.1
US-10-291-583-92

Query Match 6.8%; Score 122.5; DB 12; Length 738;
Best Local Similarity 23.0%; Pred. No. 0.0055;
Matches 78; Conservative 45; Mismatches 155; Indels 61; Gaps 19;

QY 18 DPTPSYGAETVNNYTRKQOYVSHDRKQOMTEBQREWSYGVGVTVNMSGQY-- 75
DB 4 DGYLPDM--LEDNLSGIREW---WDLKPGAPKPRANQKODDGRGLVDPGYKYLGPFG 58
QY 76 -----PTNRLEAFASDEDR-FKNELKNG-RP--RSGETRAEFGRVAKESFDEKQFQRA 126
DB 59 LDKGEPVNAADAAAEHDKAYDOQLKAGDNPIRYNHADAFQORLDE--DTSFGNIG 115
QY 127 REVASVNNALNADHSAVLDNKKELANGDALRNEDASPTVYALNTPSFKERNNG 186
DB 116 RAYFOAKKRVLEPL-----GLVEBGAKTAPGKKRPVBPSPQSPDSSTGIGKK---- 163
QY 187 NHDPNRKAVITSKFFMSGQDRSSADKRYGDPAPFAPAGTGLVDMGRDNIPRSPFS 246
DB 164 GHQPARKRL-----NFGQTSSESVDPPIGPAPGSGLSGSG--TMAAGGAPMADNN 216
QY 247 PG-----EGFVNDYGMFGAQTADADKTW--TGNNHY--APNGLAM--HYV 291
DB 217 EGADGVSSSGNWHCDSTWLDGRLVITTSRT-WALPTVNNHLYQJISNGTSGSTINDTY 275
QY 292 ESKFRWMSGYSDPDRCGAVITLTPKSNMTAPDKYQGM 330
DB 276 FGISTPW--GYEDFNR--FHCHSPRDMQRL--INNWN 307
```

```
RESULT 11
US-10-291-583-93
; Sequence 93, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 93
; LENGTH: 738
; TYPE: PRT
; ORGANISM: capsid protein of AAV serotype, clone 43.12
US-10-291-583-93

Query Match 6.8%; Score 122.5; DB 12; Length 738;
Best Local Similarity 23.0%; Pred. No. 0.0055;
Matches 78; Conservative 45; Mismatches 155; Indels 61; Gaps 19;

QY 18 DPTPSYGAETVNNYTRKQOYVSHDRKQOMTEBQREWSYGVGVTVNMSGQY-- 75
DB 4 DGYLPDM--LEDNLSGIREW---WDLKPGAPKPRANQKODDGRGLVDPGYKYLGPFG 58
QY 76 -----PTNRLEAFASDEDR-FKNELKNG-RP--RSGETRAEFGRVAKESFDEKQFQRA 126
DB 59 LDKGEPVNAADAAAEHDKAYDOQLKAGDNPIRYNHADAFQORLDE--DTSFGNIG 115
QY 127 REVASVNNALNADHSAVLDNKKELANGDALRNEDASPTVYALNTPSFKERNNG 186
DB 116 RAYFOAKKRVLEPL-----GLVEBGAKTAPGKKRPVBPSPQSPDSSTGIGKK---- 163
QY 187 NHDPNRKAVITSKFFMSGQDRSSADKRYGDPAPFAPAGTGLVDMGRDNIPRSPFS 246
DB 164 GHQPARKRL-----NFGQTSSESVDPPIGPAPGSGLSGSG--TMAAGGAPMADNN 216
QY 247 PG-----EGFVNDYGMFGAQTADADKTW--TGNNHY--APNGLAM--HYV 291
DB 217 EGADGVSSSGNWHCDSTWLDGRLVITTSRT-WALPTVNNHLYQJISNGTSGSTINDTY 275
QY 292 ESKFRWMSGYSDPDRCGAVITLTPKSNMTAPDKYQGM 330
DB 276 FGISTPW--GYEDFNR--FHCHSPRDMQRL--INNWN 307

RESULT 12
US-10-291-583-94
; Sequence 94, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
```

PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 94
LENGTH: 738
TYPE: PRT
ORGANISM: capsid protein of AAV serotype, clone 43.5
US-10-291-583-94

Query Match 6.8%; Score 122.5; DB 12; Length 738;
Best Local Similarity 23.0%; Pred. No. 0.0055;
Matches 78; Conservative 45; Mismatches 155; Indels 61; Gaps 19;

QY 18 DPYPSYGAETVNNYIRKQOVYSHRDKQOMTEBQREWLSCGVGTWVNSGQY-- 75
DB 4 DGYLPDM--LEDNLSEGRW---WDLKPAKPKANQKQDDRGVLPEYKYLGPFG 58
QY 76 -----PTNRLAFASDEDR-FKNEIKNG-RP--RSGETRAEFERVAKESTDEKGFQQA 126
DB 59 LDKGEPVNADAAALAEHDKAYDQKAGDNPYLRYNHADAERQERLQE---DTSFGGNIG 115
QY 127 REVASVNNRLENAHDESAVLDNLKKELANGNDALRNEDARSPYSALRNPSPKERN 186
DB 116 RAVFOAKKRVLEPL-----GLVESGAKTAPAKKRPVPSPPSSSTGIGKK--- 163
QY 187 NHDPSSMKAVIYSKHFWSGQDRSSADKRYGPDAPFAPAGTGLVMSRDNIPRSPTS 246
DB 164 GHQAPARKL-----NFGQIGDSSEVPDPPIGEPAPGSLSSG--TMAAGGAPVADNN 216
QY 247 PG-----EGFVNPYGFPGAQOTADADKTW---THGNHYH--APVGSIGAM--HYT 291
DB 217 EGADGVSSSGNWHCDSTWLGDRVITTSRT-WALPTNNHLYKOISSTGSGTMDNTY 275
QY 292 ESKFRNWSGSDPFDGAVITFIPKSWNTAPDKVKGW 330
DB 276 FGYSSTPW--GYFPDNR--FHCHFSRDMQRL---INNWM 307

RESULT 13

US-10-291-583-2
Sequence 2, Application US/10291583
Publication No. US20030138772A1

GENERAL INFORMATION:
APPLICANT: Gao, Guangping
APPLICANT: Wilson, James M.
APPLICANT: Alvirra, Mauricio
TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
FILE REFERENCE: UPR-02735USA
CURRENT APPLICATION NUMBER: US/10/291,583
PRIOR FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 737
TYPE: PRT
ORGANISM: capsid protein of adeno-associated virus serotype 7
US-10-291-583-2

Query Match 6.7%; Score 122; DB 12; Length 737;
Best Local Similarity 22.8%; Pred. No. 0.0061;
Matches 76; Conservative 45; Mismatches 161; Indels 52; Gaps 16;

QY 18 DPYPSYGAETVNNYIRKQOVYSHRDKQOMTEBQREWLSCGVGTWVNSGQY-- 75
DB 4 DGYLPDM--LEDNLSEGRW---WDLKPAKPKANQKQDDRGVLPEYKYLGPFG 58
QY 76 -----PTNRLAFASDEDR-FKNEIKNG-RP--RSGETRAEFERVAKESTDEKGFQQA 126
DB 59 LDKGEPVNADAAALAEHDKAYDQKAGDNPYLRYNHADAERQERLQE---DTSFGGNIG 115
QY 127 REVASVNNRLENAHDESAVLDNLKKELANGNDALRNEDARSPYSALRNPSPKERN 186
DB 116 RAVFOAKKRVLEPL-----GLVESGAKTAPAKKRPVPSPPSSSTGIGKK--- 163
QY 187 NHDPSSMKAVIYSKHFWSGQDRSSADKRYGPDAPFAPAGTGLVMS--RDRNP 241
DB 164 GQQPAKRL-----NFGQIGDSSEVPDPPIGEPAPGSLSSG--TMAAGGAPVADNN 218
QY 242 RSPSPGEGFVNDYGVFGAQTADADKTW---THGNHYHAPVGSIGAHVYESKERN 298
DB 219 ADGVGASGNWHCDSTWLGDRVITTSRT-WALPTNNHLYKOISSTGSGTMDNTY 277
QY 299 SE--GYSDFDRCVAVITFIPKSWNTAPDKVKGW 330
DB 278 STPWGYFPDNR--FHCHFSRDMQRL---INNWM 306

RESULT 14

US-10-291-583-99
Sequence 99, Application US/10291583
Publication No. US20030138772A1

GENERAL INFORMATION:
APPLICANT: Gao, Guangping
APPLICANT: Wilson, James M.
APPLICANT: Alvirra, Mauricio
TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
FILE REFERENCE: UPR-02735USA
CURRENT APPLICATION NUMBER: US/10/291,583
PRIOR FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 99
LENGTH: 736
TYPE: PRT
ORGANISM: capsid protein of AAV serotype, clone 43.20
US-10-291-583-99

Query Match 6.5%; Score 117; DB 12; Length 736;
Best Local Similarity 22.3%; Pred. No. 0.0197;
Matches 77; Conservative 46; Mismatches 147; Indels 76; Gaps 21;

QY 18 DPYPSYGAETVNNYIRKQOVYSHRDKQOMTEBQREWLSCGVGTWVNSGQY-- 75
DB 4 DGYLPDM--LEDNLSEGRW---WDLKPAKPKANQKQDDRGVLPEYKYLGPFG 58
QY 76 -----PTNRLAFASDEDR-FKNEIKNG-RP--RSGETRAEFERVAKESTDEKGFQQA 126
DB 59 LDKGEPVNADAAALAEHDKAYDQKAGDNPYLRYNHADAERQERLQE---DTSFGGNIG 115
QY 127 REVASVNNRLENAHDESAVLDNLKKELANGNDALRNEDARSPYSALRNPSPKERN 184

Search completed: September 22, 2003, 12:54:23
Job time : 64 secs

```
Db      116 RAVPQAKRVLPLGLVEBGAKTAPGKKRL-----VEQSRQEDSS 156
QY      185 GG---NHDPGRKAVITYSKHFWSGQDRSSADRRKYGDPAFRPAP-GTGLVDMSEDRN 239
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157 SGIGKTGGQPAKKL-----NFGQTGDSSEVPDPQLGEPPA---APGLGPNVTASGG 208
QY      240 IPRSPSPG-----EGFVNPYGVFGAQTADADKTV--THGNHYH--APNGSLGA 287
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      209 APMAONNEGADGVNSGWNHCDSTWLDGRVITTSRT-WALPTYNHLYKQISNGTSG 267
QY      288 M---HYESKFRMSEGSDFDRCAGVITTFPKSNMTAPDKYQGW 330
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      268 STNDYTFGYSTPW--GYDFPNR--FHCHFSRDMQRL---INNHW 306
```

```
RESULT 15
US-10-291-583-72
; Sequence 72, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identif
; FILE REFERENCE: UPN-0273USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIORITY FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72
; LENGTH: 737
; TYPE: PRT
; ORGANISM: capsid protein of AAV serotype, clone 3.3bVPI
US-10-291-583-72
```

Query Match 6.5%; Score 117; DB 12; Length 737;
Best Local Similarity 22.5%; Pred. No. 0.019;
Matches 75; Conservative 45; Mismatches 162; Indels 52; Gaps 16;

```
QY      18 DEYRPSYGRATVNNYIRKQOVYSHRDKKQOMTEQREWLSTYGVYTWVNSGY-- 75
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      4 DGYLDPW--LEDNLSGIREW--WDLKPGAPKPKANQKQDNGRLVPGYKYLGPFG 58
QY      76 -----PTRLAFASFDEDR-FKNELNG-RP--RSGETRAFEGRVAKESFDEKGFORA 126
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 LDKGPFVAADAAALHDKAKYDQQLNAGDNPLYLRNHNADAFQERLOE---DTSPGNLG 115
QY      127 KEVAVSWMRALFNHADSAYLDNLKKEJLNGNDALRNEDARSPPYSALANTPSFKERNG 186
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      116 RAVPQAKRVLPL-----GLVEBGAKTAPAKKRPVEPSPQSPDSSTGIGK--- 163
QY      187 NHDPGRKAVITYSKHFWSGQDRSSADRRKYGDPAFRPAPGTGLVDMSEDRN 241
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      164 GQGPARKL-----NFGQTGDSSEVPDPQLGEPPA---APGLGPNVTASGG 218
QY      242 RSPSPSPGEGFVNPYGVFGAQTADADKTV--THGNHYHAPNGSLGAMHYASKFRNW 298
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      219 ADGVANASGWNHCDSTWLDGRVITTSRT-WALPTYNHLYEQISEFTAGSTNNITYFGY 277
QY      299 SE--GYSDPDRCAGVITTFPKSNMTAPDKYQGW 330
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      278 STNDYTFGYSTPW--FHCHFSRDMQRL---INNHW 306
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:42:35 ; Search time 19 Seconds

(without alignments)
1675.360 Million cell updates/sec

Title: US-09-884-948-1

Sequence: 1 DSDDRVTTPAEPLDMPDPY.....ITPIKSWTAPDKVQKQMP 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r.76:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1811	100.0	406 2 JC2089	protein-glutamine
2	1802	99.5	332 2 JC7310	protein-glutamine
3	114.5	6.3	1217 2 S52714	sericinB - silkw
4	106.5	5.9	918 2 B82486	protease VCA0223
5	105.5	5.8	1888 2 T14273	zinc finger protei
6	105	5.8	325 2 T83308	hypothetical prote
7	105	5.8	920 2 T52426	dynamn-like prote
8	105	5.8	1032 2 H96619	protein T30E16.17
9	104	5.7	575 1 S03745	beta-amyase (BC 3
10	102	5.6	219 2 B90045	hypothetical prote
11	102	5.6	504 2 T31784	hypothetical prote
12	101.5	5.6	383 2 A8457	acetylflavin resista
13	100.5	5.5	926 2 B37271	A-alpha Y 3 protei
14	100	5.5	517 1 A39038	1-calcesonin, norma
15	99	5.5	2273 2 T146477	calcium channel B
16	99	5.5	2424 2 T4480	calcium channel B
17	98.5	5.4	1507 2 B47328	natural killer cel
18	98.5	5.4	2924 2 T18378	variant-specific s
19	98	5.4	350 2 S25525	outer membrane por
20	98	5.4	547 1 A56575	puff-specific nucl
21	98	5.4	1403 1 A47328	natural killer cel
22	98	5.4	3938 2 T42761	Bassoon protein -
23	97.5	5.4	625 2 A34615	profilaggrin - rat
24	97	5.4	572 2 S18732	anticoagulant, 64k -
25	97	5.4	591 2 D89783	RGN-containing 11p
26	96.5	5.3	1097 2 T31504	hypothetical prote
27	96	5.3	350 2 AF0543	hypothetical prote
28	96	5.3	350 2 AF0543	outer membrane por
29	96	5.3	992 2 B86237	protein F14N23.17

30	96	5.3	1344 2 T42637	hypothetical prote
31	96	5.3	1883 2 T13944	chromodomain helix
32	95	5.2	608 2 T32823	hypothetical prote
33	95	5.2	1253 1 A44400	myosin heavy chain
34	95	5.2	1313 2 A48467	myosin heavy chain
35	94.5	5.2	483 2 T19720	hypothetical prote
36	94	5.2	462 2 H64145	hypothetical prote
37	93.5	5.2	396 2 S13251	tropomn T - fruit
38	93.5	5.2	1261 2 G83162	respiratory nitrat
39	93.5	5.2	1722 1 T78879	retinoblastoma bin
40	93.5	5.2	1993 2 AF1450	probable peptidogl
41	93.5	5.2	2441 2 D71623	erythrocyte membra
42	93	5.1	632 2 E75057	peptidase PAB1418
43	92.5	5.1	506 2 G90000	hypothetical prote
44	92.5	5.1	570 1 S50933	myb-related protei
45	92.5	5.1	707 1 S69781	outer membrane pro

ALIGNMENTS

RESULT 1

JC2089
Protein-glutamine gamma-glutamyltransferase (BC 2.3.2.13) precursor - Streptococcus
N/Alternate names: glutamyl-peptide-amine gamma-glutamyltransferase; transglutaminase
C/Species: Streptococcus sp.
C/Date: 14-Jul-1994 #sequence revision 14-Jul-1994 #text change 07-May-1999
C/Accession: JC2089; JC2090; A46730
R/Washizu, K.; Ando, K.; Koike, S.; Hirose, S.; Matsuda, A.; Takai, H.; Motoki, M.;
Biocst. Biotechnol. Biochem. 58, 82-87, 1994
A/Title: Molecular cloning of the gene for microbial transglutaminase from Streptococci
A/Reference number: JC2089; MUID:94162748; PMID:7765334
A/Accession: JC2089
A/Molecule type: DNA
A/Residues: 1-406 <WAS>
A/Experimental source: strain S-8112
R/Takehana, S.; Washizu, K.; Ando, K.; Koike, S.; Takeuchi, K.; Matsui, H.; Motoki, M.;
Biocst. Biotechnol. Biochem. 58, 88-92, 1994
A/Title: Chemical synthesis of the gene for microbial transglutaminase from Streptococci
A/Reference number: JC2090; MUID:94162749; PMID:7765335
A/Accession: JC2090
A/Molecule type: DNA
A/Residues: 76-406 <TAK>
R/Kanaji, T.; Ozaki, H.; Takao, T.; Kawajiri, H.; Ide, H.; Motoki, M.; Shimomishi, Y.;
J. Biol. Chem. 268, 11565-11572, 1993
A/Title: Primary structure of microbial transglutaminase from Streptococcus sp. st
A/Reference number: A46730; MUID:93280110; PMID:8099353
A/Accession: A46730
A/Status: preliminary
A/Molecule type: protein
A/Residues: 76-406 <KAN>
A/Experimental source: s-8112
A/Note: sequence extracted from NCBI backbone (NCBI:133222)
C/Comment: This enzyme catalyzes an acyl transfer reaction between a gamma-carboxamide
lar crosslinking of certain proteins by gamma-glutamyl-epsilon-lysine side chain bridges
C/superfamily: protein-glutamine gamma-glutamyltransferase
C/Keywords: aminocyclotransferase; calcium; coagulation; heterotrimer; homodimer
F/1-16/Domain: signal sequence #status predicted <Sig>
F/19-75/Domain: propeptide #status predicted <Pro>
F/76-406/Product: protein-glutamine gamma-glutamyltransferase #status predicted <Mat>

Query Match	100.0%	Score 1811	DB 2	Length 406
Best Local Similarity	100.0%	Pred. No. 2.1e-128		
Matches 331	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	DSDDRVTTPAEPLDMPDPYPSYGAETVNNYIRKQVYSHRDGRKQMTBQRMTL	60	
DB	76	DSDDRVTTPAEPLDMPDPYPSYGAETVNNYIRKQVYSHRDGRKQMTBQRMTL	135	
QY	61	SYGCVGVTVNNSGQYPTNRLAFASPDDEPRKELKNGSPRGEGTRAFERVAKESFDEE	120	
DB	136	SYGCVGVTVNNSGQYPTNRLAFASPDDEPRKELKNGSPRGEGTRAFERVAKESFDEE	195	

QY 121 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYSALNRTSPF 180
 |||
 Db 196 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYSALNRTSPF 255
 |||
 QY 181 KERNGNNDPSSRMKAVIYSHGFMSGODRSSADKRYKDPDAPRPAPGGLVDMARDNI 240
 |||
 Db 256 KERNGNNDPSSRMKAVIYSHGFMSGODRSSADKRYKDPDAPRPAPGGLVDMARDNI 315
 |||
 QY 241 PRSPTPSGEGFVNFYDMFGAOTFADADKTWVTGNNHYHAPNGSLGAMHYVSKFRMNS 300
 |||
 Db 316 PRSPTPSGEGFVNFYDMFGAOTFADADKTWVTGNNHYHAPNGSLGAMHYVSKFRMNS 375
 |||
 QY 301 GYSDFDRGAVYITFIPIKSNNTAPDKYKQGP 331
 |||
 Db 376 GYSDFDRGAVYITFIPIKSNNTAPDKYKQGP 406
 |||

RESULT 2

UC7310
 Protein: glutamine gamma-glutamyltransferase (EC 2.3.2.13) - Escherichia coli
 N: Alternate names: microbial transglutaminase
 C: Species: Escherichia coli
 C: Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 17-Nov-2000
 C: Accession: J07310
 R: Yokoyama, K.; Nakamura, N.; Seguro, K.; Kubota, K.
 Biosci. Biotechnol. Biochem. 64, 1263-1270, 2000
 A: Title: Overproduction of microbial transglutaminase in Escherichia coli, in vitro refc
 A: Reference number: J07310
 A: Accession: J07310
 A: Molecule type: DNA
 A: Residues: 1-332 <YOK>
 A: Experimental source: strain JM109
 A: Note: Cross-reference
 C: Comment: This enzyme belongs to a family of enzymes that catalyzes acyl transfer betwe
 in the formation of epsilon-(gamma-glutamyl) lysine cross-linkages. This enzyme is invc
 C: Genes: mtg
 A: Gene: mtg
 C: Superfamily: Protein-glutamine gamma-glutamyltransferase
 C: Keywords: aminocyltransferase

Query Match 99.5%; Score 1802; DB 2; Length 332;
 Best Local Similarity 99.7%; Pred. No. 7.5e-128;
 Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSDRVTPPAEPLDRMPDPRPSYGAEIVANNYIRKQGYSHRDGRKQMTTEOREMT 60
 |||
 Db 2 DSDRVTPPAEPLDRMPDPRPSYGAEIVANNYIRKQGYSHRDGRKQMTTEOREMT 61
 |||
 QY 61 SYGCVGVTWVNSGQYPTNRLAFASPDDEPFKHEIKNGRPSSGTRAEFFGRVAKSSPDEE 120
 |||
 Db 62 SYGCVGVTWVNSGQYPTNRLAFASPDDEPFKHEIKNGRPSSGTRAEFFGRVAKSSPDEE 121
 |||
 QY 121 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYSALNRTSPF 180
 |||
 Db 122 KGFQARREVASVNNRALLENHDSAYLNDLKKELANGNDALRNEDARSFFYSALNRTSPF 181
 |||
 QY 181 KERNGNNDPSSRMKAVIYSHGFMSGODRSSADKRYKDPDAPRPAPGGLVDMARDNI 240
 |||
 Db 182 KERNGNNDPSSRMKAVIYSHGFMSGODRSSADKRYKDPDAPRPAPGGLVDMARDNI 241
 |||
 QY 241 PRSPTPSGEGFVNFYDMFGAOTFADADKTWVTGNNHYHAPNGSLGAMHYVSKFRMNS 300
 |||
 Db 242 PRSPTPSGEGFVNFYDMFGAOTFADADKTWVTGNNHYHAPNGSLGAMHYVSKFRMNS 301
 |||
 QY 301 GYSDFDRGAVYITFIPIKSNNTAPDKYKQGP 331
 |||
 Db 302 GYSDFDRGAVYITFIPIKSNNTAPDKYKQGP 332
 |||

RESULT 3
 S52714
 sericin1B - silkworm
 C: Species: Bombyx mori (silkworm)

C: Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
 C: Accession: S52714
 R: Garrel, A.A.; Deleage, G.G.; Prudhomme, J.J.
 Submitted to the EMBL Data Library, March 1995
 A: Description: Structure and organisation of the Bombyx mori sericin I gene and of t
 A: Reference number: S52714
 A: Accession: S52714
 A: Status: preliminary
 A: Molecule type: mRNA
 A: Residues: 1-1217 <GAR>
 A: Cross-references: EMBL:248802; NID: g755699; PID: g755700

Query Match 6.3%; Score 114.5; DB 2; Length 1217;
 Best Local Similarity 20.4%; Pred. No. 1.8;
 Matches 60; Conservative 47; Mismatches 136; Indels 51; Gaps 11;

QY 33 NTRKQGVYSHRDGKQMTTEORE--EMLSYGCYGVTVWNSGQ-----YPTNR 79
 |||
 Db 124 NYVSDGQAVAASSDARDENRSAQYQAQANWADSGYGVADRSASRRRQANYGDKD 183
 |||
 QY 80 LAFSPDEDFPKHEIKNGRPREGTRAEFFEGRVAKSPDEBKGFQARREVASVNNRALLEN 139
 |||
 Db 184 ITAASKDPSADSSRSN-----AYNRRSD-----GSSAGLSDRSASS 223
 |||
 QY 140 AHDSAYLNDLKKELANGNDALRNEDARSFFYSALNRTSPFERNGNHDPS-----RMK 194
 |||
 Db 224 SKNNNVHYFTKDSI--GGQAKSSRSFSGSDAYVNSSPDGSYVAGTRDSTSTNKKAS 281
 |||
 QY 195 AVIYS--KHFMSGODRSSADKRYKDPDAPRPAPGGLVDMARDNI PRSPTPSGGFV 252
 |||
 Db 282 STIYADDDQIRAAINDSSSKQKQSSAQISGG- KGTSV--BSKROYSDNRKRKSDAYV 338
 |||
 QY 253 NFDYGMFG--AOTFADADKTWVTGNNHYHAPNGSLGAMHYVSKFRMNSGYS 304
 |||
 Db 333 GRD-----GTAIYENKSEKTSQSTNYADONSVRSDSAASQTSKSIDRGYSD 388
 |||

RESULT 4

E82486
 proteinase VCA023 (imported) - Vibrio cholerae (strain N16961 serogroup O1)
 C: Species: Vibrio cholerae
 C: Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C: Accession: E82486
 R: Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Seller
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A: Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A: Reference number: A82035; M01D:20406833; PMID:10952301
 A: Accession: E82486
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-918 <HEI>
 A: Cross-references: GB:AE004362; GB:AE003853; NID: g9657611; PID: AA96135.1; GSPDB: G
 A: Experimental source: serogroup O1, strain N16961; biotype El Tor
 C: Genes:
 A: Gene: VCA023
 A: Map position: 2
 C: Superfamily: Bacillus thuringiensis immune inhibitor A

Query Match 5.9%; Score 106.5; DB 2; Length 918;
 Best Local Similarity 21.2%; Pred. No. 5.1;
 Matches 65; Conservative 52; Mismatches 113; Indels 77; Gaps 16;

QY 42 YSHR--DGRKQMTTEOREMTLSYGVTVWNSGQYPTNLT--AFASFED--RFXHEIKNGR 98
 |||
 Db 446 YSNRQDLKRMKRP-----LTIPASQQLTRFKMFPTEIDKDYIARVLINGR 493
 |||
 QY 99 PRSGETRA--EFEGRVAKESFDEBKGFQAR--EVASVNNRALLENHND-----ESA 145
 |||
 Db 494 PLAGNITTMDDPFKSGLVPAISGQSDGWDVDAQDLSNAMGQYVELAFYLLDGGSLAMEGL 553
 |||
 QY 146 YLDLKKELANGNDALRNEDARSFFYSALNRTSPF--ERNNGNNDPSSRMKAVIYSHGF 203
 |||

```

Db      554 YVDDLRLEVDGNQOTLLDNAMECTSSP-----AFQCFETKNGCFH-----ANHY 596
QY      204 SGQDRSSS-----AKKRYGDPDAFRPAPGTGLVMSNDRIIPSPSPGEGVAVNDVG 257
Db      559 LLIQMSHNDVDOGLAMKSPGQLMSEFGILLVWVSEADVAVK--HPGE-----G 647
QY      256 WFGQTEADADKTWTGHNHAPNGSLGAVHYESKFRNMEGYSDFPRGAVITFIPK 317
Db      648 WLGV-VPADQNALVWEKGT-----EVAQGRFPVRAATPSLDDQAP--LKLVA 692
QY      318 SWNRAPD 324
Db      693 DGNLTLED 699

```

RESULT 5

```

T14273
Zinc finger protein 106 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14273
R:Zuberi, A.R.; Christenson, G.J.; Mendoza, L.M.; Shastri, N.; Roopenian, D.C.
Immunity 9, 697-698, 1998
A:Title: Positional cloning and molecular characterization of an immunodominant cytotoxic
A:Reference number: Z17953; WUID:99060924; PMID:9846490
A:Accession: T14273
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-188 <ZUB>
A:Cross-references: EMBL:AF060246; NID:93372656; PID:93372657; PIDN:ABD04329.1
C:Genetics:
A:Note: Zfp106

```

Query Match 5.8%; Score 105.5; DB 2; Length 1888;

Best Local Similarity 19.6%; Pred. No. 15; Matches 68; Conservative 37; Mismatches 101; Indels 141; Gaps 17;

```

QY      8 PPAELDRMPDFRPSYGRALTVANNYIRKMQCYTHRDGRKQOMTEOREMLSYGCVG 67
Db      109 PPSNSQEVNSDDDDQPMQRREDRIPIQDBESYSQPRHRG-----PQRDW----- 154
QY      68 TWVNSGQYPTNRLAFSPDEDFKMLKN--GRP-----SGFTTA----- 106
Db      155 KWEKCG-----FVSTKNSFPRLNSGSPGSSVHKATGSSSTWFLHNSNGG 206
QY      107 -----EFGRAKESFDEE--KGFQRAEIVASVNNRALENHDESAVLDNLKEL 154
Db      207 WHSNNQWVDMVNYGTGRNSWSEGTGF-----PSWHNN----- 242
QY      155 ANGDALRNEDASPRYSALRNT-----PSYKERNGNHDPSEM-----X 194
Db      243 SGN-----KSSVRETSNWNVNGDGFQCGQNRPNVQXEDMTKNNKSNK 291
QY      195 AVITSKH--FMSGODRSSASDKRYGDPDARPRPGTGLVMSRRNRNPRSPGSPG 252
Db      292 PSKTSQERCKWQKQDDPKAKTR--SPEEGY-----ASDTFSEGLD 331
QY      253 NFDYQWFGAQTEADADKTWTGHNHAPNGSLGAVHYESKFRNMEGYSDFPRGAVITFIPK 317
Db      332 ERFNF-----EQRESQTKQTDTAASKINGKGTGA-----RDFPRMT 369

```

RESULT 6

```

T38308
Hypothetical protein SPAC23H3.15c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Oct-2000
C:Accession: T38308
R:Stelton, V.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
Submitted to the EMBL Data Library, September 1997
A:Reference number: Z21785
A:Accession: T38308

```

```

A:Molecule type: DNA
A:Residues: 1-325 <SKE>
A:Cross-references: EMBL:Z99163; PIDN:CA16245.1; GSPDB:GN000666; SPDB:SPAC23H3.15c
A:Experimental source: strain 972h-; cosmid C23H3
R:Stelton, V.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Wood, V.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z21737
A:Accession: T38381
A:Molecule type: DNA
A:Residues: 1-87 <DEPV>
A:Cross-references: EMBL:Z98944; PIDN:CA11599.1; GSPDB:GN000666; SPDB:SPAC23H1.01c
A:Experimental source: strain 972h-; cosmid c25H1
C:Genetics:
A:Gene: SPAC23H3.15c; SPDB:SPAC23H1.01c
A:Map position: 1

```

Query Match 5.8%; Score 105; DB 2; Length 325;

Best Local Similarity 23.0%; Pred. No. 1.7; Matches 64; Conservative 33; Mismatches 85; Indels 96; Gaps 16;

```

QY      71 NSGQYPTNRLAFSPDEDFRKN-----ELANGRPRSGETRAFEGRVAKSPDE 120
Db      32 SSSGAPQNNPPTSTTSFIRNSRRANDMGTDGSDPYAGTS-----DTK 79
QY      121 KGF-----QPARVASV--NRALENHDESAVLDNLKELANGDAL---RNDARSPF 170
Db      80 KGFNSVESRKKQSDVRGDTYSRRHDSYSSN--KYSTGNDYSYSGGRNED----- 132
QY      171 YSLKNTPFKRNNGN--DPSRM--KAVIYGRFWS-----GDRSSADKRX 216
Db      133 YST-----SGGSYTPDPSPTDDTASVQSQYNQSRRTTGCGYGRDYQSYPDT 182
QY      217 YGDPDAFRPAPGTGLVMSNDRIIPSPSPGEGVAVNDVGMPFAQTEADADKTWTGHN 276
Db      183 YG-----SRKATPSDITVGGG-----YDSSSSHHGSGHTEHRGGS 222
QY      277 H--YHAPNGSLGAVHYESKFRNMEGYS--DPRGAY 310
Db      223 YGNDNTANTRGAVS-----SAGYSGEGYKGT 251

```

RESULT 7

```

T52426
Dynammin-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C:Accession: T52426
R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.
Gene 239, 309-316, 1999
A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic reg
A:Reference number: Z25171
A:Accession: T52426
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-920 <KAT>
A:Cross-references: EMBL:AB028467; PIDN:BA88111.1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: CFI
A:Map position: 1

```

Query Match 5.8%; Score 105; DB 2; Length 920;

Best Local Similarity 24.6%; Pred. No. 6.6; Matches 59; Conservative 27; Mismatches 104; Indels 50; Gaps 11;

```

QY      26 RAEVTV--NNYRKMQCYSHRDG-----RKQOMTEQREWLSTYG 63
Db      684 KASMDKXEMWINKQYIQARFGQYSSAMQSLSBGLDKMTRKVDDEBLRMSGE 743
QY      64 CVGVTVNSGQYPTN-----RLAFASFDEDFRKNLKGPRPSGSTRAF--EGRVAYE 115
Db      744 VRGVEAVVNSIAAVPRAVVLQVEKSKEDMLNQLVYSISAIQNERISLIEDQNVKR 803

```

QY 116 SPDEKGFQARREYASVMNALENADH-----SAYLDNKKELA---NGNDALNEDAR 167
Db 804 RRDR---YQKQSSILSLTQTQL-SHDNRMAAASWSDNGTSSSRTTGGSS--GEPMW 857
QY 168 SPFYALNTPSFERNNGNHDPSRKAVIYSKHPWGGQDRSSADKKRYGDPDAFRPAP 227
Db 858 NAFNMAAGPDSLKRYSGGH--SRR---YSDPAQNGEDSSGGSSSRRTTPNRLPPAP 911

RESULT 8

H96619
Protein T30E16.17 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C/Accession: H96619
R/Theologos, A.; Ecker, J.R.; Palm, C.D.; Federici, N.A.; Kaul, S.; White, O.; Alonso,
Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Llorca, J.S.; Maltli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzman, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: H96619
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1092 <STO>
A/Cross-references: GB:AE005173; NID:g8778745; PIDN:AAF79753.1; GSPDB:GN00141
C/Genetics:
A:Gene: T30E16.17
A:Map position: 1

Query Match

Best Local Similarity 24.6%; Score 105; DB 2; Length 1092;
Matches 59; Conservative 27; Mismatches 104; Indels 50; Gaps 11;

QY 26 RAETV--NNYIRKQGVYSHRDG-----RKQMTBQREWLSYG 63
Db 856 KAESWVDKRWINKLQVIGARQGVGASMSKQSLSEGSLDKVRKFPVDEELHWGQ 915
QY 64 CYGVTVWVNSGQYPTN-----RLAFSPDEDFKYLELKNKPPRGGERAF--EGRVAK 115
Db 916 VGVYEAVLNSLAANVKAVALQVEKSKEDMLQVSSISAIIGNERISLIQEDQNTKR 975
QY 116 SPDEKGFQARREYASVMNALENADH-----SAYLDNKKELA---NGNDALNEDAR 167
Db 976 RRDR---YQKQSSILSLTQTQL-SHDNRMAAASWSDNGTSSSRTTGGSS--GEPMW 1029
QY 168 SPFYALNTPSFERNNGNHDPSRKAVIYSKHPWGGQDRSSADKKRYGDPDAFRPAP 227
Db 1030 NAFNMAAGPDSLKRYSGGH--SRR---YSDPAQNGEDSSGGSSSRRTTPNRLPPAP 1083

RESULT 9

S03745
beta-amylase (EC 3.2.1.2) precursor - Bacillus circulans
C/Species: Bacillus circulans
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: S03745
R/Siggen, K.W.
Mol. Microbiol. 1, 86-91, 1987
A/Title: Molecular cloning and characterization of the beta-amylase gene from Bacillus c
A/Reference number: S03745; MUID:88260890; PMID:2455212
A/Accession: S03745
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-575 <SIGL>
A/Cross-references: GB:Y00523; NID:g39407; PIDN:CAA68578.1; PID:g39408
C/Function:
A/Description: catalyzes the hydrolysis of 1,4-glycosidic linkages of starch, removing m

C/Superfamily: beta-amylase
C/Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>
F:37-575/Product: beta-amylase #status predicted <MAT>

Query Match

Best Local Similarity 22.9%; Score 104; DB 1; Length 575;
Matches 76; Conservative 37; Mismatches 103; Indels 116; Gaps 19;

QY 24 YGRAEVYNNYIRKQGVYSHRDGKQMTBQREWLSYGCVYMW-----NSG 73
Db 80 WQYESAGDQF--DMSYTYTYADTVKQ-----AGLKNVPISTHRCGNVG 124
QY 74 Q-----YPTNRLAFSPDEDFKYLELKNKPPRG-----ETRAEFERVAKSPDEKGFQ 124
Db 125 DDQNTPLPSLWMEKSGADEMQKDE-----SGVYNNLSLSPFSGVKGQ--YDE----- 171
QY 125 RAEEYASVMNALENADHSAVLNKKELANGNDALRNEDASPPYSALNTPSFERN 184
Db 172 ---LYASF-----KQNSAYKDWIPIKYLSGP-----SGELRYESYYPAA 209
QY 185 GGNHDPSPRKAVIYSKHPWGGQDRSSADKKRYGDPDAFRPAPGTGLVMSRDNIPSP 244
Db 210 GMSY-PARGKFPYVTE---TASAPRTMTTKYISDKINANGVILTSMSQ-----ISP 260
QY 245 TSPGSGF---VNPDIYG-----WF-----GAQTEADADKXTWTH-----GNA 277
Db 261 PTDSDGFYTGQYNTITYGDFLSWYQVLENLGVIGAAHKNVDFVFCVRLGAKISGII 320
QY 278 YHAPNGSL--GANEVYESKFRMSGEGYDPR 307
Db 321 WQNNPSPHPSAER-----AGGYDYNR 343

RESULT 10

B90045
hypothetical protein [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: B90045
R/Kiroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sano, T.; Inoue, R.; Kaito, C.; Sekimizu,
C. Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.
Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: B90045
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-219 <KDR>
A/Cross-references: GB:BA000018; PID:g13702382; PIDN:BA843523.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetics:
A:Gene: SA221

Query Match

Best Local Similarity 20.8%; Score 102; DB 2; Length 219;
Matches 46; Conservative 42; Mismatches 73; Indels 60; Gaps 11;

QY 18 DVPSPY-----GRAEVYNNYIRK-----QQVYSHRDGKQMTBQRE-- 58
Db 5 DKRDSFOYDNQONHRRQSEDSYQOYAKDPEHEPFRYNGDYREQILIEEEKS 64
QY 59 -----WLSYGVGVTVWVNSGQYPTNRLAFSPDEDFKYLELKNKPPRGGERAFERV- 112
Db 65 RSKKWL-YIILAILIIVAFVTR--ALANDSDKXSDPKYSQYKKQYENG--DQIN 120
QY 113 -----AKSPDEKGFQARREYASVMNALENADHSA-----YLDNKKELANG 157
Db 121 QGVNNAKENT---KNNQKTDIDIKNLQMOIDNKKQGRONKADSKLTQFYQOQINK-LTEA 176
QY 158 NDALRNEDASPPYSALNTPS-----FERNNG 186

Db 177 NNALKNASQKLESMLNDINTKEDSIKSLBSPFXDNGG 217

RESULT 11

hypothetical protein F13H6.4 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000

C/Accession: T31784

R/Jones, K.; Wohlschlag, P.

submitted to the EMBL Data Library, July 1997

A/Description: The sequence of C. elegans cosmid F13H6.

A/Reference number: Z21085

A/Accession: T31784

A/Status: preliminary; translated from GB/EMBL/DDAJ

A/Molecule type: DNA

A/Residues: 1-504 <JUN>

A/Cross-References: EMBL:AF016437; PIRN:AA85885.1; GSPDB:GN00023; CESP:F13H6.4

A/Experimental source: strain Bristol N2; clone F13H6

C/Genetics:

A/Map position: 5

A/Intons: 39/1; 63/3; 92/3; 123/2; 205/1; 264/3; 393/1; 436/2

C/Superfamily: cholinesterase; cholinesterase homology

Query Match

Best Local Similarity 5.6%; Score 102; DB 2; Length 504;

Matches 69; Conservative 44; Mismatches 144; Indels 80; Gaps 13;

Db 28 ETVYNNYIRKQGVYSHDRKQOMTEBQEMLSYSG---CIGYTWVNSGQYPTNNLARA 83

Db 149 DVCNGNFGMLDQTLA-----LKNVQKHISFSGDPCVCTVFGSAGASATDLSIS 200

Db 84 SPDEPRKMLKNGRPSGRTAEPEGRVAK-----ESFDEKQFQARAVASV---NW 134

Db 201 PHSDDLFQRI-----PISGAYCEPRLRTSKSQAKIFRFAPFKFTGDSSTILLEWYKN 256

Db 135 FALNNAHDESAYLDNKKELANGNDALRNEDAR-----SPFYSAIRNTPSFKRNNGNHD 189

Db 257 QSSFTLSD--LRKEAPKKQMTGVDEYEGVLAAMNPEFSPADAGIALFPKQVGNLTAN 314

Db 190 PSRMKAVIYSKHFWSGODRSSGADRK-----GGDDAFRPAFGTGLVMSRDRNTPRSP 244

Db 315 PEBMHKLFYKXY-BEVDKSDSANKKRLCEAFG----- 348

Db 245 TSPEGEVNDPYGMF-GAQTEDADKVTWTHG-NHYHAPNGSL-----GANHYEERK 294

Db 349 -----LGFNLGVFQSAKSAKYGNDVFLYSEFYHSDGPGMWKDLFPFASNGHELR 401

Db 295 FRWNSGYSDFDRGAVITFIPKSNTPAPDKVKQGP 331

Db 402 YLLGEGFYSKFDATKEELEVEKTTLPFSNFAYKGNP 438

RESULT 12

acetylflavin resistance protein E [imported] - *Brucella melitensis* (strain 16M)

C/Species: *Brucella melitensis*

C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C/Accession: AH3457

R/DelVecchio, V.G.; Kaputal, V.; Redkar, R.U.; Patra, G.; Mujar, C.; Los, T.; Ivanova,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A/Reference number: AD3252; PMID:11756688

A/Accession: AH3457

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-383 <KUR>

A/Cross-References: GB:AE008917; PIRN:AA52827.1; PID:G17993667; GSPDB:GN00190

A/Experimental source: strain 16M

C/Genetics:

A/Genes: EMBL1646

A/Map position: 1

Query Match 5.6%; Score 101.5; DB 2; Length 383;

Best Local Similarity 26.18; Pred. No. 3.9;

Matches 29; Conservative 21; Mismatches 32; Indels 29; Gaps 3;

Db 63 GCYGTWVNSGQYPTNNLAPASFD-----DRFKNELKNGRPSGRTAEFEGR 111

Db 189 GIGGILPNNAGNYTACTGSTARLDDSTYLDIWIWPERFAQIVQGPULASTA-FGE 247

Db 112 VAVESFDESKGFQARAVASVNNALNNAHDESAYLDNKKELANGNDALR 162

Db 248 IYKGRIN-----AVDNMLDEASPTLHVRLEVYNNADRLR 281

RESULT 13

B37271
A-alpha Y 3 protein - bracket fungus (*Schizophyllum commune*)

C/Species: *Schizophyllum commune*

C/Date: 07-Feb-1992 #sequence_revision 07-Feb-1992 #text_change 05-Dec-1997

C/Accession: B37271

R/Julich, R.C.

submitted to the Protein Sequence Database, October 1991

A/Reference number: A37271

A/Accession: B37271

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-926 <ULU>

C/Superfamily: unassigned homeobox proteins; homeobox homology

C/Keywords: DNA binding; homeobox; nucleus; transcription regulation

F/148-204/domain: homeobox homology <HGX>

Query Match

Best Local Similarity 5.5%; Score 100.5; DB 2; Length 926;

Matches 71; Conservative 36; Mismatches 122; Indels 117; Gaps 15;

Db 9 PAELDRPDPYRPSYGAETVNNYIRKQGVYSHDRKQOMTEBQEMLSYSGVGT 68

Db 601 PAEPYIR-FDDPAFPVALAEKRAKRRARKEKKQAEKARK----- 640

Db 69 WNVNSGQYPTNNLAPASFDPRKQELKNGRPSGRTAEFEGRVAKSPDEKQFQAR 126

Db 641 -----EKKRAKRAKQKKDKRQKGLPR-SPEITLDS-----SRAS 678

Db 129 VASVNNALNNAHDESAYLDNKKELANGNDALRNEDKSPFYSAIRNTPSFKRNNGN 188

Db 679 VTSDAGATSRKSRTRKRPDSASVAS-----ARTP-SLSTSS-RSSGTS 724

Db 189 DRS--RMK---AVIYSKHFWSGODRESSAD---KRTGDPDA----- 222

Db 725 MPATPMMESLPPVASDNFVLTGDKVITWTPELMAQLGEGDDAGLDEPNQSGFSPDL 784

Db 223 -FRPAPGTGLVMSRDRNIPRSPTPSGEGVNFQDYGFG--AQTEADADKVTWTHGNHY 279

Db 785 IFSSCNDGALGMDTNDNME-----LGDLSPTQLSFDMMNTTSMDS 828

Db 280 A-ENGSLGMEVYSKFRNWSGYSDFDRGAVITFIPKSNTPAPDKVKQGP 324

Db 829 TQPAASFDSSSETSSMDPNW-----LTPQCANTPAD 859

RESULT 14

A39038
1-caldesmon, nonmuscle - chicken

N/Alternate names: caldesmon, nonmuscle

C/Species: *Gallus gallus* (chicken)

C/Date: 31-Jul-1991 #sequence_revision 27-Jun-1994 #text_change 22-Jun-1999

C/Accession: A39038

R/Hayashi, K.; Fujio, Y.; Kato, I.; Sobue, K.

J. Biol. Chem. 266, 355-361, 1991

A/Title: Structural and functional relationships between h- and l-caloesmons.

A/Reference number: A39038; MUID:91093148; PMID:1824698

A/Accession: A39038

A/Molecule type: mRNA
A/Residues: 1-517 <HAV>
A/Cross-references: GB:M60620; GB:M38015; NID:g212242; PIDN:AAA48936.1; PID:g212243
A/Experimental source: brain
C/Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C/Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon
C/Superfamily: caldesmon
C/Keywords: actin binding; calmodulin binding; phosphoprotein
P:342,427/Binding site: phosphate (Ser) (covalent) #status predicted
P:433,456/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match

5.5%; Score 100; DB 1; Length 517;
Best Local Similarity 18.7%; Pred. No. 7.4;
Matches 60; Conservative 41; Mismatches 116; Indels 104; Gaps 12;

```
QY 28 ETVANNYIR-KWQVYSHRDKQKMTTEORRMWSYGVGVYTWNSQVYPTNRILAFASFD 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 151 ETVYKSYQRNNRQ-DGEEKKEKESSEK-----PREVPT-----E 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 EDRFK-NEIKNGRPSGRTAEFGVAKESPEDEKGFQARREVASVNNRLENA----- 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 189 ENGVKMKVKEKAPKE-BMKVMD--RKGVPQKQNGERRELTTPKLSYENAFGRSN 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 141 -----HDSAYLNLKKELANGDALRNEPARSPYSALRNTPSFKE 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 LKGAANAEKSEKLEKQKQEAVALDELKRRERKTLERESQKKQKEFAEKIRESEE 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 RINGNHDPKMKAVIYSKHPMSGDRSSADK-----RKYDDPD 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 KKMKEIERERRAEAEKROKVPEDGVSEKKPKCFSPKGSLSKTEERAEPLNKAOKS 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 222 AFRPAPGTGLVMSRDR-----NTPRSPTS-----PGEGVNPDYGMFGAQT 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 GMPKATTAIVWSKIDSLRLEQTSAYVGNKAKAPAPASDLIPVAGVRNI----- 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 EADADKTVTGTHGNHHPNGS 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 416 ----KSMWEKGNVSSPGT 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

146477
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C/Accession: 146477; 146478
R/Mori, Y.; Friedlich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Bosse, E.; Hofman
Nature 350, 398-402, 1991
A/Title: Primary structure and functional expression from complementary DNA of a brain c
A/Reference number: 146477; MUID:91187110; PMID:1849233
A/Accession: 146477
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2273 <MOR>
A/Cross-references: EMBL:X57476; NID:g1522; PIDN:CAA40714.1; PID:g1523
A/Accession: 146478
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1856, 'H', 1858, 'K', 1860-1862, 'SL', 1865-1866, 'VIS', 1870-1876, 'K', 1878-1879, '
A/Cross-references: EMBL:X57688; NID:g1524; PIDN:CAA40871.1; PID:g1525
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match

5.5%; Score 99; DB 2; Length 2273;
Best Local Similarity 26.5%; Pred. No. 60;
Matches 43; Conservative 18; Mismatches 61; Indels 40; Gaps 7;

```
QY 99 PRSGTRAEFGVAKESPEDEKGFQARREVASVNNRLEVAHDSAYLNLKKELANGN 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 828 PQENNNNTNKSRLAEPFTVDRLGQRAEDFLKQARHDPADPSAH----- 875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 DALRNEDARSPYSA-----LNTPSFKE-----RNGNHPSPRKRAVIYSKHPMSGQDR 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 876 -AAAGLDARRPWAGSGAEILSRGPGYGRSDHQAREGGLEPPG-----FWEGE-- 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

QY 209 SSSADKKRYGDDPAPRPAPGTGLVMSRDRN-IPRSPTSPGE 249
DB 923 ---AERGKADPPR-RHARQGVGSGSGSRSGSFRITGTADGE 960

Search completed: September 22, 2003, 12:45:29
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:41:14; Search time 14 Seconds

(without alignments)
1111,846 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDDRVTPPAEPDPRDPY.....ITPDKSWNPADPKXQGM 331

Scoring table: ELUCUM62

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1811	100.0	331 1	TGL STRESS
2	108	6.0	1382 1	IP3A_HUMAN
3	104	5.7	575 1	AMTB_BACI
4	100.5	5.5	926 1	MAY3_SCHCO
5	100	5.5	744 1	JPH3_MOUSE
6	99.5	5.5	943 1	ARS2_MOUSE
7	99	5.5	2424 1	CCAA_RABIT
8	98.5	5.4	1453 1	NKCR_MOUSE
9	98	5.4	350 1	PHOE_SALTY
10	98	5.4	547 1	EX42_DROME
11	98	5.4	1462 1	NKCR_HUMAN
12	96	5.3	350 1	PHOE_SALTY
13	96	5.3	497 1	AMYP_STECA
14	96	5.3	536 1	SKIP_HUMAN
15	96	5.3	1344 1	IP3A_MOUSE
16	96	5.3	2468 1	MABP_HUMAN
17	95	5.2	572 1	LMD1_HUMAN
18	94	5.2	462 1	YFHD_HAEN
19	93.5	5.2	396 1	TET_DROME
20	93.5	5.2	1722 1	RBB2_HUMAN
21	93	5.1	482 1	UZR2_HUMAN
22	92.5	5.1	633 1	MTH_TERTH
23	92.5	5.1	790 1	KTF3_MOUSE
24	92	5.1	660 1	JPH1_MOUSE
25	92	5.1	1101 1	DIA2_HUMAN
26	92	5.1	2459 1	MABP_RAT
27	91.5	5.1	381 1	NCAP_CVCAI
28	91.5	5.1	383 1	ONS2_SALTY
29	91	5.0	350 1	RS40_ARATH
30	91	5.0	535 1	YGH1_CABEL
31	91	5.0	636 1	GYSB_THEMA
32	91	5.0	1536 1	SIN3_YEAST
33	90.5	5.0	351 1	PHOE_KLEPN

34	90.5	5.0	367 1	OMPC_ECOLI	P06986	eschertichia
35	90.5	5.0	632 1	YGA9_SCHPO	O43070	schizosach
36	90.5	5.0	980 1	BOH1_YEAST	P38041	saccharomyc
37	90.5	5.0	1131 1	MOG1_CABEL	P24041	caenorhabdi
38	90	5.0	351 1	PHOE_CITFR	O01605	citrobacter
39	90	5.0	611 1	IF4B_HUMAN	P23588	homo sapien
40	89.5	4.9	537 1	ARP_ELARA	P04931	plasmodium
41	89.5	4.9	935 1	KINH_SYNRA	O43053	syncephalas
42	89.5	4.9	3421 1	TEGU_HYVEB	P28955	equine herp
43	89	4.9	559 1	QOAI_ARATH	O91593	arabidopsis
44	89	4.9	977 1	DLP1_HUMAN	O14490	homo sapien
45	88.5	4.9	532 1	CCB1_DROAC	O44220	drosophila

ALIGNMENTS

[Handwritten signature]

RESULT 1	TGL STRESS	STANDARD	PRT: 331 AA.
AC	P81453		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13)		
DE	(Transglutaminase) (Tcase)		
OS	Streptococcus sp. (strain s-8112)		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Streptococcaceae; Streptococcaceae; Streptococcus.		
OX	NCEB1_TaxID=86037;		
EN	[1]		
RP	SEQUENCE.		
RX	MEDLINE=33280110; Pubmed=8099353;		
RA	Kenji T., Ozaki H., Takeo T., Kawajiri H., Ide H., Motoki M.,		
RA	Shimomishi Y.,		
RT	Primary structure of microbial transglutaminase from		
RT	Streptococcus sp. strain s-8112.		
RL	J. Biol. Chem. 268:11565-11572(1993).		
CC	- FUNCTION: Catalyzes the cross-linking of proteins and the		
CC	configuration of polyamines to proteins.		
CC	- CATALYTIC ACTIVITY: Protein glutamine + alkylamine = protein N(5)-		
CC	alkylglutamine + NH(3)		
CC	- MASS SPECTROMETRY: MW=37869.2; MW_ERR=8; METHOD=Electrospray.		
CC	- BIOTECHNOLOGY: Sold under the name Activa TG by Ajinomoto. It has		
CC	the ability to crosslink protein molecules present in food without		
CC	the use of salt or binders. Used to improve some of the physical		
CC	properties such as firmness, elasticity and moisture retention of		
CC	food such as meat, poultry and seafood.		
DR	PDB: 1J14; 27-AUG-02.		
DR	Transferrase; Acyltransferase; 3D-structure.		
FT	ACT SITE 64		
FT	POTENTIAL.		
SO	SEQUENCE 331 AA; 37862 MW; 5992363A3B2C1FA CRC64;		
Query Match	100.0%; Score 1811; DB 1; Length 331;		
Best Local Similarity	Pred. No. 1.5e-129; Mismatches 0; Gaps 0;		
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 DSDDRVTPPAEPDPRDPYRPSYGRALTYVNNYIRKWOQVYSHRDGKQKQMTEDQEWL 60		
DB	1 DSDDRVTPPAEPDPRDPYRPSYGRALTYVNNYIRKWOQVYSHRDGKQKQMTEDQEWL 60		
QY	61 SYGCVGVTVNNSQVYTNLAFASPDPRFNNLKNGRPSGETFAEFGVAESFDES 120		
DB	61 SYGCVGVTVNNSQVYTNLAFASPDPRFNNLKNGRPSGETFAEFGVAESFDES 120		
QY	121 KQGFARAEVAVSVNRRALAEVDESAVLDNLKELANGNDALNEDARSPYSALNTPSP 180		
DB	121 KQGFARAEVAVSVNRRALAEVDESAVLDNLKELANGNDALNEDARSPYSALNTPSP 180		
QY	181 KERNGCNHDSRRKAVITYSKHFWSGQDRSSADRKRYGPDPAFRAPPTGIVMSGRDRI 240		
DB	181 KERNGCNHDSRRKAVITYSKHFWSGQDRSSADRKRYGPDPAFRAPPTGIVMSGRDRI 240		

QY 241 PRSPSPGEGFVNFYDYGAGQIADADKVTWTHGNHYPNGSIGAMHYEKSFEFNMSE 300
 DB 241 PRSPSPGEGFVNFYDYGAGQIADADKVTWTHGNHYPNGSIGAMHYEKSFEFNMSE 300
 QY 301 GYSPDFRGAVYITTPKSMWNTAPDKYQGWMP 331
 DB 301 GYSPDFRGAVYITTPKSMWNTAPDKYQGWMP 331
 RESULT 2
 IPI3A_HUMAN STANDARD; PRT; 1382 AA.
 AC Q14152; 000653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Eukaryotic translation initiation factor 3 subunit 10 (eIF-3 theta)
 DE (eIF3 p167) (eIF3 p160) (eIF3 p185) (eIF3a).
 GN EIF3S10 OR KIAA0139.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96127530; PubMed=8590280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. IV.
 RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 2:167-174(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97234683; PubMed=9150439;
 RA Scholler J.K., Kanner S.B.;
 RT "The human p167 gene encodes a unique structural protein that contains
 RT centrosomal A homology and associates with a multicomponent complex.";
 RL DNA Cell Biol. 16:515-531(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97207269; PubMed=9054404;
 RA Johnson K.R., Weirick W.C., Zolli W.L., Zhu Y.;
 RT "Identification of cDNA clones for the large subunit of eukaryotic
 RT translation initiation factor 3. Comparison of homologues from human,
 RT Nicotiana tabacum, Caenorhabditis elegans, and Saccharomyces
 RT cerevisiae.";
 RL J. Biol. Chem. 272:7106-7113(1997).
 RL J. Biol. Chem. 272:7106-7113(1997).
 CC -I- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
 CC METHIONYL-TRNAI AND MRNA.
 CC -I- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -I- DOMAIN: CONTAINS 1 PCI DOMAIN.
 CC -I- PTM: PHOSPHORYLATED.
 CC -I- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D50929; BA09488.1; -
 DR EMBL; U58046; AAB41584.1; -
 DR EMBL; U78314; AAB80695.1; -
 DR Genew; HGNC:3271; EIF3S10.
 DR GK; Q14152; -
 DR MIM; 602039; -
 DR GO; GO:0005852; Eukaryotic translation initiation factor 3. . . TAS.
 DR GO; GO:0006446; P:regulation of translational initiation; TAS.
 DR InterPro; IPR000717; PCL.

DR Pfam; PF01399; PCL; 1.
 DR SMART; SM0088; PINT; 1.
 KW Initiation factor; Protein biosynthesis; Repeat; Phosphorylation.
 FT DOMAIN 925 1172 25 x 10 AA TANDEN REPEAT OF D-IDEI-D-R-
 FT (GP) - (PS) - (RW) - R - (GN) - (AM).
 SQ SEQUENCE 1382 AA; 166568 MM; 485C01B28D67BBA CRC64;
 Query Match 6.0%; Score 108; DB 1; Length 1382;
 Best Local Similarity 21.5%; Pred. No. 2.5; Indels 98; Gaps 12;
 Matches 68; Conservative 33; Mismatches 118;
 QY 39 QQYSHRGRKQQTTEQREWLISYGVVYWNISQYPTVRLAFASDFDRFKMLNGR 98
 DB 785 EERHNRLEERKQKEERR-----ITYRKEEERQRRAEQMLERRERRARRA 835
 QY 99 PPSGTRAEFEGRYAKSPFEKGFORAEVAVSWNRRLNADDAIYDNTKGLANGN 158
 DB 836 KKEEELR-EYQERVKLEVEYERKQRELEIEERRRR-----EERRRIGD 880
 QY 159 DALRNDAR-----SPFSALRNTSPFKE-RNGNHDPDRMKAVYKHPWS 204
 DB 881 SLSKRDRGMDRDSBGTWKKPEADSEWRNRPPEKWRGGERD----- 925
 QY 205 GQDRSSGADK---RKYDDP---AFRP---APGTGVNNSDRNIPSPSPGEGVYN 253
 DB 926 -BDRSRDRDESRPRRLDDEDEPRSLRPDDRVPARGDDRGPRGPEDEFSRGADD 984
 QY 254 FDGVFGAGQTS-----ADADKVTWTHGNHYPNGSIGAMHYEKSFEFNMSEGSDDPR 307
 DB 985 DRSFNRNDDPRPRRLADDRGNWHEADDPRPRRL-----DDDR 1026
 QY 308 GAVYITTPKSMWNTAPD 324
 DB 1027 G-----SWRTADE 1034
 RESULT 3
 AMYB_BACCI STANDARD; PRT; 575 AA.
 ID AMYB_BACCI
 AC P06547;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Beta-amylase precursor (EC 3.2.1.2) (1,4-alpha-D-glucan
 DE maltohydrolase).
 DE Maltotetraose.
 OS Bacillus circulans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 11033;
 RX MEDLINE=88260890; PubMed=2455212;
 RA Siggen K.W.;
 RT "Molecular cloning and characterization of the beta-amylase gene from
 RT Bacillus circulans.";
 RL Mol. Microbiol. 1:86-91(1987).
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
 CC polysaccharides so as to remove successive maltose units from the
 CC non-reducing ends of the chains.
 CC -I- SUBUNIT: Monomer.
 CC -I- SIMILARITY: BELONGS TO FAMILY 14 OF GLYCOSYL HYDROLASES
 CC (BETA-AMYLASES).
 CC -----
 DR EMBL; Y00523; CAA68578.1; -
 DR PIR; S03745; S03745.

CC	use by non-profit institutions as long as its content is in no commercial
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M97180; AAB01370.1; -.
DR	PIR; B37271; B37271.
DR	InterPro; IPR001356; Homeobox.
DR	Pfam; PF04611; Alaphay_MDB; 1.
DR	Pfam; PF00046; Homeobox; 1.
DR	ProDom; PD00010; Homeobox; 1.
DR	SMART; SM00389; HOX; 1.
DR	PROSITE; PS00027; HOMEBOX 1; 1.
DR	PROSITE; PSS0071; HOMEBOX 2; 1.
DR	Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
KW	Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT	DNA_BIND 147 206
FT	DOMAIN 620 660 ARG/LYS-RICH (BASIC).
FT	DOMAIN 677 724 SER-RICH.
SQ	SEQUENCE 926 AA; 10183 MW; 1866317422D3E2E8 CRC64;
Query Match	5.5%; Score 100.5; DB 1; Length 926;
Best Local Similarity	20.5%; Pred. No.5.6; Indels 117; Gaps 15;
Matches	71; Conservative 36; Mismatches 122;
Qy	9 PABPLDMEDPFRPSYSGAEITVNNYTRKQQVYSHRGQRQQMTBEBREMLSYGCVYT 68
Dd	601 PAEPDYR-PDDPAPFVALAKRAARAKRKKKQQAERK- 640
Qy	69 WMSGGYPFNRLAFASFPEDPFKNELKMGSPRASGETRLEFGGRPAKSFPDEKGORAAE 128
Dd	641 -----EKRRARRKAQAQKRCKQRAGLPRR-SPTLDS-----SRAS 678
Qy	129 VASYMFALENADHDSAYLDLKLKELANGNDALRNEDARSFYSALENTSPFKERGNGH 188
Dd	679 VTSDASASTRKRTSRKRKRDSSASVAS -----AKT---SLSTGS--RRSGTS 724
Qy	189 DPS--RWK--AVITYSKHFWSGDRESSAD----KKRYGDPDA----- 222
Dd	725 MPATPPNMNLSLPVVASDNFVLGTDKDVWTPEIMQLTGEDDASGLDEPMOSGEFSDDL 784
Qy	223 -FRARPDTGLVMSNDRIIPRSPPSBEGBVPDVGWG-ACQEAADATVTWTHGNHYH 279
Dd	785 IFSSCNDGALGMNTADVNP- ----LGLSDTLQSLFDMMWTSMDSL 828
Qy	280 A-PWGSICAHVYESKFRNWSEGYSDPDRCGVYTFIPKSNTPAD 324
Dd	829 TOPAASPSSSETSSMDEN- ----LLPQCANTAPD 859
RESULT 5	
JPH3_MOUSE	
ID	JPH3_MOUSE STANDARD; PRT; 744 AA.
AC	O9ET77; O9EQ22; Q9ET77; Q9EQ22; Rel. 41; Created!
DT	28-FEB-2003 (Rel. 41; Last sequence update)
RC	28-FEB-2003 (Rel. 41; Last sequence update)
PT	15-SEP-2003 (Rel. 42; Last annotation update)
DE	Unctophilin 3 (uncTopophilin type 3) (CP-3).
GN	JPH3 OR JP3.
OS	Mus musculus (mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX	NCEI_TaxID=10099;
RN	[1]
RP	SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC	STRAIN=C57BL/6J; and 129; TISSUE=Brain;
RX	MEDLINE=2040329; PubMed=10949023;
RA	Takeshima H., Komazaki S., Nishi M., Ino M., Kangawa K.;
RT	"UncTopophilins: a novel family of junctional membrane complex
RL	proteins";
RU	Mol. Cell 6:11-22(2000).
RN	[2]
RP	FUNCTION, AND TISSUE SPECIFICITY.
RM	MEDLINE=21903756; PubMed=11906164;

RA Nishi W., Hashimoto K., Kuriyama K., Komazaki S., Kano M., Shibata S.,
 RA Takeshima H.;
 RT "Motor discoordination in mutant mice lacking junctophilin type 3";
 RL Biochem. Biophys. Res. Commun. 292:318-324(2002).
 CC -1- FUNCTION: Contributes to the stabilization of the junctional
 CC membrane complexes, which are common to excitable cells and
 CC mediate cross-talk between cell surface and intracellular ion
 CC channels. Probably acts by anchoring the plasma membrane and
 CC endoplasmic reticulum (By similarity). May play an active role in
 CC certain neurons involved in motor coordination.
 CC -1- SUBCELLULAR LOCATION: Type IV membrane protein. Localized
 CC predominantly on the plasma membrane. The transmembrane domain is
 CC anchored in endoplasmic reticulum membrane, while the N-terminal
 CC part associates with the plasma membrane.
 CC -1- TISSUE SPECIFICITY: Specifically expressed in brain. Expressed in
 CC certain populations of neurons but not in glial cells. In
 CC cerebellar sections, it is highly expressed in Purkinje cells,
 CC while it is weakly expressed in granular cells.
 CC -1- DOMAIN: The KORN (membrane occupation and recognition nexus)
 CC repeats contribute to the plasma membrane binding, possibly by
 CC interacting with phospholipids (By similarity).
 CC -1- MISCELLANEOUS: JPH3 deficient mice are viable and fertile but have
 CC defects in balance/motor coordination tasks.
 CC -1- SIMILARITY: BELONGS TO THE JUNCTOPHILIN FAMILY.
 CC -1- SIMILARITY: Contains 8 KORN repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AB024449; BAB12046.1; -
 DR EMBL; AB024450; BAB20320.1; -
 DR MGD; MGI:1891497; Jph3.
 DR InterPro; IPR003409; KORN.
 DR Pfam; PF02493; KORN; 8.
 DR SMART; SM00698; KORN; 6.
 KW Endoplasmic reticulum; Transmembrane; Repeat.
 KW CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1 723 ANCHOR FOR TYPE IV MEMBRANE PROTEIN
 FT TRANSMEM 724 744
 FT FT
 FT REPEAT 15 37 (POTENTIAL).
 FT REPEAT 39 60 MORN 1.
 FT REPEAT 61 82 MORN 3.
 FT REPEAT 83 105 MORN 4.
 FT REPEAT 107 129 MORN 5.
 FT REPEAT 130 152 MORN 6.
 FT REPEAT 153 175 MORN 7.
 FT REPEAT 176 198 MORN 8.
 FT DOMAIN 4 143 GLY-RICH.
 FT DOMAIN 144 366 ALA-RICH.
 FT SEQUENCE 744 AA; 81229 MW; 3D72AE06A6FDA914 CRC64;
 Query Match 5.5%; Score 100; DB 1; Length 744;
 Best Local Similarity 22.0%; Pred. No. 4,7;
 Matches 63; Conservative 35; Mismatches 112; Indels 76; Gaps 14;
 QY 23 SYORAE-----TYNNYIRKMOQ-----YSHRDQRKQMTREBRLS-- 61
 DB 265 SLSEAEALAEVEDIDDTTTEYVGEWKDKRSGFGVSGSDGLKT---EGEWSNR 319
 QY 62 ---YGCV---GVTVWSGQYPTNRALAFASPFDEFRKNELKNGFRSGETRAFEGRVAK 114
 DB 320 RHVGCGMTPEPDGTR--EGKKKQWVLY-----SGKKNLIPLRASK 358
 QY 115 ESFDEKGFQRRARFASVNRALFNHDSAYINDLKEIANGNDALRMNDASPFYSAL 174
 DB 359 IREKVDRAVEALERRAATIAQKQELIASRISH-SPAKKAALTAAQKQGEARITARIK 417
 QY 175 RNTPSFKRNGNHPDRMKAVYS--KFWNGQDPS--SSADRKRYGDPDAPRAPPG 230

DB 418 EFSPFORENG-----LEYQRPKHQKSCDIEVLTSTPPLQGSPELYR--KGT 466
 QY 231 LVDMSRD-----RNTPRSPTS--PGEQFNPFVYFGAGTADADKT 270
 DB 467 PSDLPDPSPLQSPASPSTPEPPAPASRTMAHFSQVSDERS 512
 RESULT 6
 RS22 DROME
 ID RS22 DROME STANDARD; PRT; 943 AA.
 AC Q9V9K7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aresenite-resistance protein 2 homolog.
 GN C97843.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoinkins R.A., Galie R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Betos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bokora D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Bocsa S., Dunkov B.C., Dunn P.,
 RA Dubin K.J., Evansglista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gottlieb J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.D., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Klumpe B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrelti A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazovic M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,
 RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 [2]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkeley;
 RX MEDLINE=22425069; PubMed=12537572;
 RA Maier S., Crosby W.A., Munger C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochownik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celinker S.B., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.T., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.B.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 CC -1- ALTERNATIVE PRODUCTS;
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=long;
 CC IsoId=09V9K7-1; Sequence=Displayed;
 CC Name=short;
 CC IsoId=09V9K7-2; Sequence=VSP_000327;
 CC -1- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003784; AAM68343.1;
 DR EMBL; AE003784; AAM68345.1;
 DR Flybase; FB95003062; CG7843.
 DR Pfam; PF04959; ARS2; 1.
 KM Hypothetical protein; Alternative splicing.
 FT VARSPLIT 47 50 Missing (in isoform short).
 FT /FTD=VSP_000327.
 SQ SEQUENCE 943 AA; 107221 MW; 0C1AF03E02E8AB08 CRC4;
 Query Match 5.5%; Score 99.5; DB 1; Length 943;
 Best Local Similarity 18.3%; Pred. No. 6.9;
 Matches 69; Conservative 43; Mismatches 144; Indels 121; Gaps 13;
 QY 10 AEPLDRPDPYRSGYGAATVNNYIKQVYSHDGRKQWTEBQEWLSTGCVGW 69
 DB 311 ADPSTGRKRPV-----VNSGEMDDDAENSAKKEIDSKD----- 351
 QY 70 VNSQYPTNLEAFASFEDERFKNEIKNGRPR-----SGETRAEFGRVAKESFDEE 120
 DB 352 --SSSKP-----EDKQMKTKYKKRKNSSDDSSSESSSSSDEKIKEXYDVE 399
 QY 121 KGPRAEFAVAVNRLBNADDESAVYDNKKEIANGNDALNDASPRYSALRTPEF 180
 DB 400 DGLFAEKTEAEKDRBARKAKQSPQPKLDEDEGENETEPKGDJDKINTYHSIDTLKS 459
 QY 181 KE-----RNGCNHDPSEK-----KAVYSEHWSGQD-----SSAEKXKGDPAFR 224
 DB 460 PEISNPRIKRTDNGDSKVEDEGKPEV-----GKQVWETETIDLDKVKQGPRAH 512
 QY 225 PAPTGTAVMSRDRNIPRSP-SPGEGFVN-----FDYGFQAQTE 264
 DB 513 RTSSIF-----RNLAISITSEIEVCNRFSGYLRVALADPVERRWRRGMITFMD 566
 QY 265 ADADKVTWTHGN-----HYAPNGSLGANHYVES----- 293
 DB 567 VNIKEITWGINNORLDCEMALIVNRLSRVPRANGITAHQVVRSDILCAKIALND 626
 QY 294 -KFRNMSGYSDFDGCA 309
 DB 627 EKFLMAEGPKDSDNSA 643
 RESULT 7
 CCAA_RABIT STANDARD; PRT; 2424 AA.
 AC P27884; P27883;
 RT 01-JUL-1993 (Rel. 26, Created)
 RT 01-JUL-1993 (Rel. 26, Last sequence update)
 RT 15-SEP-2003 (Rel. 42, Last annotation update)
 RT Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
 RT channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel
 RT 1) (B1).

GN CACNA1A OR CACNA1A4 OR CAC4 OR CACN3.
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CX NCBI_Taxid=9986;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=9118710; Pubmed=184923;
 RA Mori Y., Friedrich T., Kim M.-S., Mikami A., Nakai J., Ruth P.,
 RA Bosse E., Hofmann F., Flockeitz V., Fumichl T., Mikoshida K.,
 RA Imoto K., Tanabe T., Numa S.;
 RT Primary structure and functional expression from complementary DNA
 RT of a brain calcium channel.";
 RL Nature 350:398-402(1991).
 RN [2]
 RP BETA-SUBUNIT BINDING DOVAIN AND MUTAGENESIS.
 RX MEDLINE=94150724; Pubmed=7509046;
 RA Pragnell M., de Waard M., Mori Y., Tanabe T., Snutch T.P.,
 RA Campbell K.P.;
 RT "Calcium channel beta-subunit binds to a conserved motif in the I-II
 RT cytoplasmic linker of the alpha 1-subunit.";
 RL Nature 368:67-70(1994).
 CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE
 CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
 CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
 CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
 CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
 CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
 CC IVA (OMEGA-A&A-IVA). THEY ARE HOWEVER INSENSITIVE TO
 CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA).
 CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 CC FORMING AND VOLTAGE-SENSITIVE GENERATE VOLTAGE-SENSITIVE CALCIUM
 CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=BI-2; Synonyms=1A-2;
 CC IsoId=P27884-1; Sequence=Displayed;
 CC Name=BI-1; Synonyms=1A-1;
 CC IsoId=P27884-2; Sequence=VSP_000879, VSP_000880;
 CC Name=CBP101; Synonyms=CBP109;
 CC IsoId=P27884-3; Sequence=VSP_000878;
 CC Name=CBP103;
 CC IsoId=P27884-4; Sequence=VSP_000877;
 CC Name=CBP107;
 CC IsoId=P27884-5; Sequence=VSP_000876;
 CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC. PURKINJE CELLS CONTAIN
 CC PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
 CC CURRENT IN CEREBELLAR GRANULE CELLS.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57477; CAA40715.1; -.

DR EMBL; X57688; CAA40872.1; -
 DR EMBL; X57476; CAA40714.1; -
 DR EMBL; X57688; CAA40871.1; -
 DR PIR; 146477; 146477.
 DR PIR; 146480; 146480.
 DR InterPro; IPR001682; Ca/Na pore.
 DR InterPro; IPR002077; Ca channel.
 DR InterPro; IPR002111; Cat channel, TrpL.
 DR InterPro; IPR005821; Ion trans.
 DR InterPro; IPR005820; M-channel nlg.
 DR InterPro; IPR005448; PoydCCAlphal.
 DR Pfam; PR00520; ion trans; 4.
 DR PRINTS; PR00167; CACHANNEL.
 DR PRINTS; PR01632; PoydCCAlphal.
 DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KW Calcium channel; Glycoprotein; Repeat; Multigene family;
 KW Calcium-binding; Phosphorylation; Alternative splicing.
 FT REPEAT 85 363
 FT REPEAT 473 717
 FT REPEAT 1240 1523
 FT REPEAT 1560 1823
 FT DOMAIN 1 98
 FT TRANSSEM 99 117
 FT TRANSSEM 118 135
 FT TRANSSEM 136 155
 FT TRANSSEM 156 167
 FT TRANSSEM 168 185
 FT TRANSSEM 186 190
 FT TRANSSEM 191 209
 FT TRANSSEM 210 228
 FT TRANSSEM 228 248
 FT TRANSSEM 249 335
 FT TRANSSEM 336 360
 FT TRANSSEM 361 487
 FT TRANSSEM 488 506
 FT TRANSSEM 507 521
 FT TRANSSEM 522 541
 FT TRANSSEM 542 549
 FT TRANSSEM 550 568
 FT TRANSSEM 569 578
 FT TRANSSEM 579 597
 FT TRANSSEM 598 616
 FT TRANSSEM 617 636
 FT TRANSSEM 637 689
 FT TRANSSEM 690 714
 FT TRANSSEM 715 1253
 FT TRANSSEM 1254 1272
 FT TRANSSEM 1273 1288
 FT TRANSSEM 1289 1308
 FT TRANSSEM 1309 1320
 FT TRANSSEM 1321 1339
 FT TRANSSEM 1340 1350
 FT TRANSSEM 1351 1369
 FT TRANSSEM 1370 1388
 FT TRANSSEM 1389 1408
 FT TRANSSEM 1409 1495
 FT TRANSSEM 1496 1520
 FT TRANSSEM 1521 1575
 FT TRANSSEM 1576 1604
 FT TRANSSEM 1605 1609
 FT TRANSSEM 1610 1629
 FT TRANSSEM 1630 1637
 FT TRANSSEM 1638 1656
 FT TRANSSEM 1657 1665
 FT TRANSSEM 1666 1684
 FT TRANSSEM 1685 1703
 FT TRANSSEM 1704 1723
 FT TRANSSEM 1724 1785
 FT TRANSSEM 1796 1820
 FT TRANSSEM 1821 2424
 FT DOMAIN 13 18
 FT DOMAIN 727 732
 FT DOMAIN 1004 1010

FT DOMAIN 1012 1017
 FT DOMAIN 2219 2227
 FT DOMAIN 2242 2246
 FT DOMAIN 2288 2297
 FT DOMAIN 2298 2301
 FT DOMAIN 2372 2377
 FT DOMAIN 2411 2416
 FT DOMAIN 383 400
 FT SITE 318 318
 FT SITE 668 668
 FT SITE 1469 1469
 FT SITE 1765 1765
 FT MOD RES 1831 1831
 FT CA BIND 1849 1860
 FT CARBOHYD 283 283
 FT CARBOHYD 1665 1665
 FT VARSPIC 772 1051
 FT VARSPIC 772 1120
 FT VARSPIC 1857 1884
 FT VARSPIC 2230 2273
 FT VARSPIC 2274 2424
 FT VARIANT 419 419
 FT VARIANT 877 877
 FT VARIANT 1104 1104
 FT VARIANT 386 386
 FT MUTAGEN 389 389
 FT MUTAGEN 392 392
 FT MUTAGEN 400 400
 SQ SEQUENCE 2424 AA; 273228 MM; F7CC4DD0AB48604 C9C64;
 Query Match 5.5%; Score 99; DB 1; Length 2424;
 Best Local Similarity 26.5%; Pred. No. 25;
 Matches 43; Conservative 18; Mismatches 61; Indels 40; Gaps 7;
 QY 99 PRSGTRAFPGGRVAKSPDEKGFRRARVAVSWNADLANADBSAYLDNKKELANGN 158
 DB 828 PGNENNNTNNSRVAEPTVDQRLGQGRADPLRQARRHRRADPSAH----- 875
 QY 159 DALRNEADRSFYA-----LRNTPSEK-----RNGNDPGRMVAIVYSKFWGQDR 208
 DB 876 -AAAGLDRRWMAQSQAISREGPRGRSHQRRBGLLEPG-----FWEGE-- 922
 QY 209 SSADKRYKQDPDAFRPAPGTGLVWSDRN-IPRSPTSPGE 249
 DB 923 ---AERKAGDPHS-RHAEQGVGSGSGSGSPRTGADGE 960
 RESULT 8
 NCR MOUSE STANDARD; PRT; 1453 AA.
 ID NCR MOUSE
 AC P30415;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DR 28-FEB-2003 (Rel. 41, Last annotation update)
 DS NK-tumor recognition protein (Natural killer cells cyclophilin-
 related protein) (NK-TR protein).
 GN NKR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9313824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells."
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=BALB/c; TISSUE=Blood;
RA Anderson S.K.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC INVOLVED IN THE FUNCTION OF NK CELLS.
CC -1- SIMILARITY: Contains 1 cyclophilin-like p1ase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04289; AAA37500.2; ALT_INIT.
DR HSSP; Q27450; 1A33.
DR MGD; MGI:97346; NCTR.
DR InterPro; IPR002130; CSA_P1ase.
DR Pfam; PF00160; pro_lisomerase; 1.
DR PROSITE; PS00170; CSA_P1ase_1; 1.
DR PROSITE; PS0072; CSA_P1ase_2; 1.
KM Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT DOMAIN 1 176 P1ase; CYCLOPHILIN-TYPE.
FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).
FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).
FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
FT DOMAIN 198 273 ARG/SER-RICH.
FT DOMAIN 468 565 ARG/SER-RICH.
FT DOMAIN 658 812 ARG/SER-RICH.
FT DOMAIN 1303 1453 ARG-SER TANDWM REPEAT-RICH.
SQ SEQUENCE 1453 AA; 163439 MW; DFL173F814B283E CRC64;
Query March 5.4%; Score 98.5; DB 1; Length 1453;
Best Local Similarity 17.7%; Pred. No. 14; Index 195; Gaps 21;
Matches 86; Conservative 63; Mismatches 152;
QY 1 DSDDRVTPP-AEFLDRMPDPYRPSYRA-----ETVNNYTRKQQ-- 40
DB 355 DDDSGSTPPHMKKEBQRL-RAYRPPSGKWKSGDKLSDPCSRMDEKSLQSRSSWSYNG 413
QY 41 VYS-----HRDG-----RKQMTEDREWLISGCVGTWYN 71
DB 414 YSSDLSTARSDGHNRKREKFKKKAKKQKQGRHRRQYKRR-----IWPDELE 466
QY 72 SGQVPTNRL-----APASPEDRFNNEKKNRPSGRTAEFGRYAKE 115
DB 467 PSRPFTRKMSGCVRRRSRRSSSSSHHSKRWMSKDDDDSSATHSRSRY--RSKSH 524
QY 116 SFDEKGFQAREVAVNRALENADESA-----YLDWLKELANGDALRNDAR 167
DB 525 SRSSRSRSSSRRAVSKSSSRSLNRSKSSSSSGSPRTSISPKRQACSEKPKYTEPPR 584
QY 168 -----SPPYALP----- 175
DB 585 PSYQNGNVLVQPAENIPVPLSDSPSPKPKQKPKWKSYSRIQEMZAKTTHLPV 644
QY 176 -----NTPSEKERNGNHDSRKKAIVYSKHPWGGDGRSSADKRYGDDPD 222
DB 645 QSTYSLTNITATVSSSSYHKR--EKPSBSDSGASYSKY--SPRSSGSSGR--SSSKS 694
QY 223 FRAPAGTGLVMSRDRNIP-----RSPT-----SPGEGVNFDPYGFAGTADAD---- 268

```

```

DB 695 SRSSRSSRYTRSRSRSLPTSRSLSPSSRSHSPKXSDGQSHSRSSSTVSVDGGR 754
QY 269 -----KTVTHGNHYHAFNGLAMEV-----YESKFRWSGQSDPD-----RGAVV 311
DB 755 AMFRNKKKSVTSHKRRNSERTLHSHKIVRGREKSHHKKYSESSSLDYSDQSHV 814
QY 312 ITFIPKSWNTAPDKVYKQ 329
DB 815 QVY-----SAPEKEXQ 826
RESULT 9
PROE_SALTY
ID PROE_SALTY STANDARD; PRT; 350 AA.
AC P30705;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PROE OR STM0320.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=93083954; PubMed=1280609;
RA Splierings G., Elders R., van Litch B., Hofstra H., Tommassen J.;
RT "Characterization of the Salmonella typhimurium phoB gene and
RT development of Salmonella-specific DNA probes.";
RL Gene 122:45-52(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Splierh J., Clifton S.W., Latreille P.,
RA Courtney L., Portwell L.S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stonking T., Nhan M.,
RA Waterson R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN POSE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68023; CA448164.1; -.
DR EMBL; AE008709; AAL19276.1; -.
DR PIR; S25525; S25525.
DR HSSP; P02932; 1PHO.
DR StyGene; SGI0291; phoB.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin Gram-ve.
DR Pfam; PF00267; Gram-ve porins; 1.
DR PRINTS; PR00182; SCOLNEIPORIN.
DR Prodom; PF000808; OMP_2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.

```

KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.

FT STGNL 1 20 OTHER MEMBRANE PORE PROTEIN E.
SQ CHAIN 21 350
SQ SEQUENCE 350 AA: 38762 MW: 9DLB3355AF59877 CRC64;

Query Match 5.4%; Score 98; DB 1; Length 350;

Best Local Similarity 20.8%; Pred. No. 2.6;

Matches 64; Conservative 35; Mismatches 116; Indels 92; Gaps 13;

24 YGRATVNNVIRKQGVSHDGRKQKQTEQREWLSTYCCVGTWNS---GQIPTN 78

34 YGVKAM-----HYSDYDSKDG-----DQSYRFFKFEKSTQINDQLTGKRWEAR 79

79 RLAFASPDDEFKKELNGSPSGETRAFFGRVAKD--SFDEKQFCRAEYASVWNR 136

80 ---FASNKAESDSQK-----TLAFMGLTKLKGISFDGRGLGALIDVETWDMF 128

137 LBNADBSAYLDNKKELANGDALNEDAPSPYSALNTSPFKERNQGNHDSRMKAV 196

129 PFGGDSAGQTDNMTKASGLATVRND---PFGIVDGLDITLQYQKNEHD--- 179

197 TSKKFNQGDSSAPKRYGPDPAFPAAGTGLVMSNDKRIKSPSPSGRGTVPDY 256

180 -VKQNGDGFCTSVSYD--FGSSD-FAVSAVYTLSPRTSEQLQRRGT----- 223

257 GWFGAQTADAKTWTGTHGNHYHAPNGSLGAMHYESKFNWSEGYSDPFGAVYTFIP 316

224 ---GDKARA-----NATGVKYA-----NDVIATFYS 248

317 KSMNTAP 323

249 ETRNMTP 255

RESULT 10

ID Bx42_DROME STANDARD; PRT; 547 AA.

AC P39736; Q9W390; 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Puff specific protein Bx42.

GN Bx42 OR CG8284.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R; TISSUE=Embryo;

RX MEDLINE=93048378; PubMed=1424966;

RA Wieland C., Mann S., von Besser H., Saumweber H.;

RT "The Drosophila nuclear protein Bx42, which is found in many puffs on

polytene chromosomes, is highly charged.";

RL Chromosoma 101:517-525(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20186006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,

RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abell J.P., Agrevari A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Bendable J., Bayraktaroglu L., Basley E.M.,

RA Beeson K.Y., Bense P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Broctler P.,

RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkoc B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Fiertera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Gang N.S., Galbert W.M., Glasser K.,

RA Glodex A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibsen C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kamsion J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasio P., Lei J., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Koshirel A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Mzy D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Musser D.R., Pacleb J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reiter K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weinsteinbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [3]

RP REVISIONS.

RC STRAIN=Berkley;

RX MEDLINE=2242606; PubMed=12537572;

RA Miste S., Crosby N.A., Murgall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Bettencourt B.R., Ceiniker S.E., de Grey A.D.N.J., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.,

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

systematic review.";

RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=2242606; PubMed=12537569;

RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,

RA George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,

RA Rubin G.M., Ceiniker S.E.,

RT "A Drosophila full-length cDNA resource.";

RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).

CC -1- FUNCTION: MAY PLAY A ROLE IN CHROMATIN STRUCTURE AND FUNCTION.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DEVELOPMENTAL STAGE: TWO TRANSCRIPTS ARE DETECTED OF SIZES 1.9 AND

2.2 KB. BOTH ARE DETECTED SOON AFTER FERTILIZATION AND SHOW

RELATIVELY CONSTANT EXPRESSION DURING THE FIRST 2/3 OF

EMBRYOGENESIS. IN 0-3 HR EMBRYOS, THE SMALLER TRANSCRIPT IS

PREDOMINANT AND THE LEVELS OF THE TWO TRANSCRIPTS ARE SOMEWHAT

REDUCED AT THE LATER STAGES OF DEVELOPMENT, BUT THEY ARE FOUND IN

APPROXIMATELY CONSTANT AMOUNTS DURING LARVAL, PUPAL AND ADULT

STAGES. THE SMALLER TRANSCRIPT IS SUSPECTED TO BE A WATERAL

TRANSCRIPT.

-1- SIMILARITY: BELONGS TO THE SWM FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL: X64536; CAA45834.1; -

EMBL: AB003446; AAF46444.1; -

EMBL: AYL1364; AAM29369.1; -

PIR: A56575; A56575.

DR FlyBase: FBgn004856; Bx42.
 DR InterPro: IPR004015; SKIP_SMW.
 DR Pfam: PF02731; SKIP_SMW; 1.
 KW Nuclear protein; DNA-binding.
 FT DOMAIN 177 343 SMW.
 FT DOMAIN 383 462 SH2-LIKE DOMAIN.
 FT DOMAIN 209 236 PRO-RICH.
 FT DOMAIN 125 155 ASP/GU-RICH (ACIDIC).
 FT DOMAIN 300 400 ASP/GU-RICH (ACIDIC).
 SQ SEQUENCE 547 AA; 61156 MW; 01399EA291C9D557 CRC64;
 Query Match 5.4%; Score 98; DB 1; Length 547;
 Best Local Similarity 22.6%; Pred. No. 4.5;
 Matches 73; Conservative 38; Mismatches 116; Indels 96; Gaps 19;
 QY 44 HHDGKQOQTEBQREMLSYGCVGWVWNSGOYPTNLA-----FAQFD 86
 DB 234 HSPSKVTV-KEQKEMKIPPCIS-NWKNAKVTTLPLKRLAAGRLQGVHNEKFAAMA 291
 QY 87 E-----DFKNELKNGPRSGETRAFEFGVAKESPDSEKGFQAREVASVNNRALENVAH 141
 DB 292 EALYIADKKAE-----AVEARSQLKELAKK--EKEKEDMLRMA-----QEAR 335
 QY 142 DSSATLDLKKELANGDALRNDARSPFYALNTSPFRNGGNDHDSRKAKVYSKH 201
 DB 336 EERRA-----GARPEAEPSGSGATGS-EVEREN----- 363
 QY 202 FMSGQDRSSADKKKYCDPPAPAP-CTGLVDSRPDRNPSS-----PTSPGEGFYNF 254
 DB 364 -----DFAERQRQRQDRRLQRAPEKRSKIQERERDLSQIAGLPAKSAKNGETLP 418
 QY 255 DVGWFGAQTEDA--DKTWTHGNHYAP--NGSIGAHVYESKFRNWSGY-SDPDR 307
 DB 419 DQRLNPTTKGMDSCYGDDEAY--NVDKPMWDSNTLGA-HYRPSKQSDSNYGDLDA 474
 QY 308 GAVVTFIP-KSNWTPADPKYKQ 329
 DB 475 IVTNTRFVFDKQFSGAKSAAAG 497
 RESULT 11
 NKCR_HUMAN STANDARD; PRT; 1462 AA.
 ID NKCR_HUMAN
 AC P30414;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NK-tumor recognition protein (Natural killer cell cyclophilin-related protein) (NK-TR protein).
 GN NKTR.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RX MEDLINE=93133824; PubMed=8421688;
 RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A., Ortaldo J.R.;
 RA "A cyclophilin-related protein involved in the function of natural killer cells.";
 RA Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
 RP REVISIONS.
 RA Anderson S.K.;
 RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
 CC INVOLVED IN THE FUNCTION OF NK CELLS.
 CC -1- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane via its N-terminus.
 CC -1- SIMILARITY: Contains 1 cyclophilin-like ppiase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: L04288; AAA5734.2; -
 DR EMBL: AF184110; AAD56402.1; -
 DR HSSP: Q27450; 1A33.
 DR Genew; HGNC:7833; NKTR.
 DR MTM; 161565; -
 DR GO: GO:0004600; F:cyclophilin; TAS.
 DR InterPro: IPR002130; CSA_Ppiase.
 DR Pfam: PF00160; ppiase; 1.
 DR PRINTS: PR00153; CSA_PPIASE.
 DR PROSITE: PS00170; CSA_PPIASE_1; 1.
 DR PROSITE: PS0072; CSA_PPIASE_2; 1.
 KW Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
 FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.
 FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).
 FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).
 FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).
 FT DOMAIN 194 244 ARG/SER-RICH.
 FT DOMAIN 466 574 ARG/SER-RICH.
 FT DOMAIN 664 814 ARG/SER-RICH.
 FT DOMAIN 1311 1348 ARG-SER TANDEN REPEAT-RICH.
 SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;
 Query Match 5.4%; Score 98; DB 1; Length 1462;
 Best Local Similarity 20.0%; Pred. No. 15;
 Matches 48; Conservative 38; Mismatches 80; Indels 74; Gaps 11;
 QY 9 PAELPDRNP-----DPRPSGAEVY-VNNIRKMQQVVS-----H 44
 DB 606 PVIPLSDSPPSRKNQKQKPKPKSTYRIQEKAKVTTLPLTQSTYLANIKETGSSSYH 665
 QY 45 RDRKQOQTEBQREMLSYGCVGWVWNSGOYPTNRLAFASFDREKNEKNGPRSGET 104
 DB 666 K--REKNSESPQSTYKX-----SDRSSESPSRSSRS 699
 QY 105 RAPEFGVAKESPDSEKGFQAREVASVNNRA--LENAHDEAYLDNKEKLANGDAL 161
 DB 700 RS-----YRSYTRSRSLASSHSRSPSSRSRKYKGDHQCRCSSSYSI 747
 QY 162 RNEDARSFYALNTPSPFRNGGNT--DPSRKAVIYKHEKWSQDPSADPKRYGD 219
 DB 748 SSDDGR---RAKRLRSSGKKNVSHKHSSESEKTLHKKV-KGRDRSCV--RKYSE 800
 RESULT 12
 PHOE_SALTI STANDARD; PRT; 350 AA.
 ID PHOE_SALTI
 AC Q56119;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Outer membrane pore protein B precursor.
 GN PHOE OR STY0365 OR T2530.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 CX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=IMSS-1;
 RA Torres A., Puente J.L., Calva E.;
 RA Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 CC [2]
 CC SEQUENCE FROM N.A.
 CC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Church C., Mungall K.L., Bentley S.D., Holden M.T.G., Seibin M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jags K.,
 RA Krog A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhimurium CT18.";
 RL Nature 413:848-852(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burtland V., Kodyolami V., Schwartz D.C., Blatner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhimurium
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC - FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
 CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
 CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
 CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
 CC SOLUTES.
 CC - SUBUNIT: Homotrimer.
 CC - SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC - SIMILARITY: BELONGS TO THE OMPC/PHO3 FAMILY OF PORES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X74595; CAA52672.1; -
 DR EMBL: AL627266; CAD08790.1; -
 DR EMBL: AE016842; AA070114.1; -
 DR PIR: S36606; S36606.
 DR HSSP: P02932; LPHO.
 DR InterPro: IPR003429; OMP_2.
 DR InterPro: IPR001702; Porin_Gram-ve.
 DR Pfam: PF00267; Gram-ve_porins; 1.
 DR PRINTS: PR00182; ECOLI_PORIN.
 DR PRODOM: PD003808; OMP_2; 1.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 1 20
 FT CONFLICT 63 63 E -> K (IN REF. 1).
 FT CONFLICT 80 80 F -> G (IN REF. 1).
 FT CONFLICT 229 229 A -> G (IN REF. 1).
 FT SEQUENCE 350 AA; 38744 MW; 89055F8CE9F0C77 CRC64;
 SO
 Query Match 5.38; Score 96; DB 1; length 350;
 Best local similarity 20.58; Pred. No. 3;
 Matches 62; Conservative 33; Mismatches 125; Indels 82; Gaps 11;

DB 184 NGDGFSGSYD---FGSD-FAVSGAYTLSDRTEDOUJQRGT-----CD 225
 QY 262 QTEADADKVTWTHGHHYHAPNGSLGAMHYVESKFNMBEGSDDRGAYITLTPKSNVT 321
 DB 226 KAEE-----WATGVKYDA-----NDIYATVYSETRMW 253
 QY 322 AP 323
 DB 254 TP 255
 RESULT 13
 AMPV STRCA STANDARD; PRT; 497 AA.
 ID AMPV STRCA
 AC P83053;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-amylase, pancreatic (EC 3.2.1.1) (1,4-alpha-D-glucan
 DE glucanohydrolase) (Pancreatic alpha-amylase) (PA).
 OS *Struthio camelus* (Ostrich).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
 OC Struthio.
 OX NCBI_TaxID=8801;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Pancreas;
 RX MEDLINE=21176845; PubMed=11281265;
 RA Kabuto S., Ogawa T., Muramoto K., Oosthuizen V., Naude R.J.;
 RT "The amino acid sequence of pancreatic alpha-amylase from the ostrich,
 RT *Struthio camelus*.";
 RL Comp. Biochem. Physiol. 127B:481-490(2000).
 RN [2]
 RP SEQUENCE OF 1-53.
 RC TISSUE=Pancreas;
 RA Oosthuizen V., Naude R.J., Oelofsen W., Muramoto K., Kamiya H.;
 RT "Ostrich pancreatic alpha-amylase: kinetic properties, amino terminal
 RT sequence and substrate structure.";
 RL Int. J. Biochem. 26:1313-1321(1994).
 CC - CARBOHYDRATE ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC - COPOLYMER: Binds a calcium ion required for its activity. In
 CC mammals it also encloses one chloride ion which activates the
 CC enzyme.
 CC - SUBCELLULAR LOCATION: Secreted. Extracellular.
 CC - SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 DR InterPro: IPR006593; Alp_amyl_cat_sub.
 DR InterPro: IPR006048; Alpha_amyl_C.
 DR InterPro: IPR006047; Alpha_amyl_cat.
 DR InterPro: IPR006046; Glyco_hydro_13.
 DR Pfam: PF00128; alpha-amylase; 1.
 DR Pfam: PF02805; alpha-amylase C; 1.
 DR PRINTS: PR00110; ALPHAMYLASE.
 DR SMART: SM00642; Amy; 1.
 DR SMART: SM00632; Amy; C; 1.
 KM Hydroxylase; Glycosidase; Carbohydrate metabolism; Calcium; Chloride;
 KW Pyridoxine carboxylic acid.
 FT MOD RES 1 1
 FT ACT_SITE 101 101 PYRROLIDONE CARBOXYLIC ACID.
 FT ACT_SITE 195 195 SUBSTRATE BINDING (BY SIMILARITY).
 FT ACT_SITE 197 197 CHLORIDE BINDING (BY SIMILARITY).
 FT ACT_SITE 233 233 BY SIMILARITY.
 FT ACT_SITE 298 298 CHLORIDE BINDING (BY SIMILARITY).
 FT ACT_SITE 299 298 CHLORIDE BINDING (BY SIMILARITY).
 FT ACT_SITE 300 300 SUBSTRATE BINDING (BY SIMILARITY).
 FT ACT_SITE 337 337 BY SIMILARITY.
 FT METAL 100 100 CHLORIDE BINDING (BY SIMILARITY).
 FT METAL 159 159 CALCIUM (BY SIMILARITY).
 FT METAL 167 167 CALCIUM (BY SIMILARITY).
 FT METAL 201 201 CALCIUM (BY SIMILARITY).

RA Polk P., Puta F., Krpejskova L., Blahuskova A., Markos A.,
RA Rabin M., Dotchin R.P.,
RT "The homolog of chromatin binding protein Bx42 identified in
RT Dictyostelium.",
RT Gene 181:229-231(1996).
CC -1- FUNCTION: INTERACTS WITH THE SKI ONCOGENE. INVOLVED IN VITAMIN D-
CC MEDIATED TRANSCRIPTION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SNW FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isdb.ch/announce>
CC or send an email to license@isdb-sib.ch).
CC
CC EMBL: U51432; AAC15912.1; -
CC EMBL: AF045184; AAC31697.1; -
CC EMBL: AC008372; AAR23125.1; -
CC EMBL: U43960; AAB4887.1; -
CC TRANSFAC: T04597; -
CC
CC MIM: 603055; -
CC DR GO: GO:0005634; C:nucleus; TAS.
CC DR GO: GO:0003713; F:transcription co-activator activity; TAS.
CC DR GO: GO:0007048; P:oncogenesis; TAS.
CC DR GO: GO:0006375; P:regulation of transcription from Pol II pro. .; TAS.
CC DR InterPro: IPR004015; SKIP_SNW.
CC DR Pfam: PF02731; SKIP_SNW; I.
CC Kew Nuclear Protein.
CC
CC FT DOMAIN 174 339 SNW.
CC FT DOMAIN 219 233 PRO-RICH.
CC FT DOMAIN 373 453 SH2-LIKE DOMAIN.
CC SO SEQUENCE 536 AA; 61494 MW; 0CC7E0D0B2CF842 CRC64;

Query Match 5.3%; Score 96; DB 1; Length 536;
Best Local Similarity 20.2%; Pred. No. 6.2; Matches 16; Gaps 14;
Matches 67; Conservative 42; Mismatches 116; Indels 106; Gaps 14;

QY 44 HSDRKKQOMTEQREWLSTVGVYTWNSQY--PTN-----RLA 81
DB 23 HSPSRKQTV-KEDQEKFKPPCLIS-MNKNAGIYIPLDKRLADRGLOTVINENAKLA 288
QY 82 PASFDEDFRNNELKNGRPSGSGTSAEEFGSVAKESFDEKFGQRAREVASM----- 133
DB 289 EALYIADRKARE-----AVEMRAQYERKCAOK--EKEHEEKLREMAOKAREPAGIK 339
QY 134 -----NHALENAHDSAYLDNTEKKELANNDALNENBARSPFSAIR 175
DB 340 THEKEDGEARREDDTIRHRRKREKQHDR-----NLSRAAPDKSKLQNNENRDISETYL 394
QY 176 NTFPFKERNGNHDP-----SRMKAV-----IYSKFFMSGQR-----SSSA 212
DB 395 GVNPAPRSTNEVQYDQLFLFQSKMGDSGFGAGEDEIYNYVDMRGKDMAQSIYPSKUL 454
QY 213 DKRYKGPDDPRAPATGTVDMKSDRIIPSPSPG-----EFTVNFVDGKFGFGQTGA 265
DB 455 DKDYGDDLAR-----IKTNREVPDPKESGSDRRQGRGREGVQFEDDPDGDKFL 505
QY 266 DADKTVTWTHGNTHAENGSLGAMHYVESKPR 296
DB 506 EEAK---QHGGSKRPDSDSRPKHEHHEGKR 533

RESULT 15
ID IF3A_MOUSE STANDARD; PRT; 1344 AA.
AC P3116; O60697; O62182;
DT 01-NOV-1991 (Rel. 20, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Buxarjodic translation initiation factor 3 subunit 10 (eIF-3-theta)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 22, 2003, 12:42:00 ; Search time 40 Seconds

(without alignments)
2135.385 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811
Sequence: 1 DSDRVTPPAEPDRMPDPY.....ITFIPKSWNTAPDKVKQGP 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mirc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriaph:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1811	100.0	376	2 Q9ZAF5	Q9ZAF5 streptomyc
2	1811	100.0	407	2 Q8KJ2	Q8KJ2 streptomyc
3	1732	95.6	334	2 Q8KJ5	Q8KJ5 streptomyc
4	1487	82.1	416	2 Q8GR90	Q8GR90 streptomyc
5	1432	79.1	411	2 Q9RIS1	Q9RIS1 streptomyc
6	1432	6.7	737	12 Q8UQ90	Q8UQ90 adeno-assoc
7	114.5	6.3	1217	5 Q17240	Q17240 bombix mori
8	113.5	6.3	738	12 Q8UQF8	Q8UQF8 adeno-assoc
9	112.5	6.2	1161	5 Q9W2X8	Q9W2X8 drosophila
10	111.5	6.2	758	10 Q8I525	Q8I525 oryza sativ
11	110	6.1	736	12 Q65311	Q65311 adeno-assoc
12	109	6.0	919	2 Q07351	Q07351 vibrio chol
13	108	6.0	401	4 Q96C72	Q96C72 homo sapien
14	108	6.0	558	4 Q9H0G5	Q9H0G5 homo sapien
15	107	5.9	736	12 Q56139	Q56139 adeno-assoc
16	106.5	5.9	736	12 Q56137	Q56137 adeno-assoc

17	106.5	5.9	736	12 Q9W2X8	Q9W2X8 drosophila
18	106.5	5.9	894	16 Q8PB66	Q8PB66 xenithomonas
19	106.5	5.9	918	16 Q9KX06	Q9KX06 vibrio chol
20	106	5.9	776	5 Q8UT08	Q8UT08 encephalit
21	105.5	5.8	839	11 Q8C235	Q8C235 mus musculu
22	105.5	5.8	865	3 Q8X0P4	Q8X0P4 neurospora
23	105.5	5.8	1888	11 Q88466	Q88466 mus musculu
24	105	5.8	325	3 P78890	P78890 schizosach
25	105	5.8	920	10 Q9SLT4	Q9SLT4 arabidopsis
26	105	5.8	921	10 Q9SLT3	Q9SLT3 arabidopsis
27	105	5.8	1092	10 Q9U055	Q9U055 arabidopsis
28	104	5.7	219	16 Q8N0Z8	Q8N0Z8 staphylococ
29	104	5.7	636	16 Q8S252	Q8S252 oryza sativ
30	103.5	5.7	1245	11 Q8R314	Q8R314 mus musculu
31	102.5	5.7	123	12 Q9PZV7	Q9PZV7 hepatitis d
32	102	5.6	219	16 Q99RJ9	Q99RJ9 staphylococ
33	102	5.6	504	5 Q16352	Q16352 caenorhadi
34	102	5.6	735	12 Q56652	Q56652 adeno-assoc
35	101.5	5.6	383	16 Q8YF76	Q8YF76 brucella me
36	101.5	5.6	383	16 Q8G2P2	Q8G2P2 brucella su
37	100.5	5.5	285	5 Q8WRP0	Q8WRP0 caenorhadi
38	100.5	5.5	467	10 Q94LH8	Q94LH8 oryza sativ
39	100.5	5.5	801	11 Q8B0Q2	Q8B0Q2 mus musculu
40	100.5	5.5	966	11 Q8C3P6	Q8C3P6 mus musculu
41	100	5.5	630	5 Q9GNN5	Q9GNN5 branchiost
42	100	5.5	730	5 Q8T867	Q8T867 dictyosteli
43	100	5.5	744	11 Q8BR43	Q8BR43 mus musculu
44	100	5.5	750	11 Q8BNM7	Q8BNM7 mus musculu
45	100	5.5	775	5 Q9Y015	Q9Y015 plasmodium

ALIGNMENTS

RESULT 1

ID Q9ZAF5 PRELIMINARY; PRT; 376 AA.

AC Q9ZAF5;
DT 01-MAY-1999 (TRMBLrel. 10, Created)
DT 01-MAY-1999 (TRMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TRMBLrel. 20, Last annotation update)
DB Transglutaminase (EC 2.3.2.13) (Fragment).
OC Streptomyces mobaraensis.
CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=35621;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSWZ;
RC MEDLINE=9053680; PubMed=9839945;
RA Pasternack R., Dorsch S., Otterbach J.T., Robenek I.R., Wolf S.,
RA Fuchsbauer H.B.;
RT "Bacterial pro-transglutaminase from Streptovorticillium mobaraense :
RT purification, characterisation and sequence of the zymogen."
RL Eur. J. Biochem. 257:570-576(1998).
DR EMBL; Y18315; CAA7128.1; -;
KW Acyltransferase; Transferase.
FT NON TER
SQ SEQUENCE 376 AA; 4245 MM; 15FE7474DE3771B9 CRC64;

Query Match 100.0%; Score 1811; DB 2; Length 376;

Best Local Similarity 100.0%; Pred. No. 1,4e-128;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DSDRVTPPAEPDRMPDPYRSGAETVNNYTRKQGVSHRGRKQMTESQREML 60
DB	46	DSDRVTPPAEPDRMPDPYRSGAETVNNYTRKQGVSHRGRKQMTESQREML 105
QY	61	SYGCVGVTVNNSGQYPTNRIAFASFDIEDRFKTELKNGRPPSGETRAEFGRVAKESFDEE 120
DB	106	SYGCVGVTVNNSGQYPTNRIAFASFDIEDRFKTELKNGRPPSGETRAEFGRVAKESFDEE 165
QY	121	KQFQARAEVVASVNNRRLNNHADSAYLDNKKKELANGNDLNNEDARSFPYSALNTPSF 180

```

Db      166 KGFQARREVAIVMNALENHDESAVLDNLKKELANGDARLNEDARSPFYALNTPSF 225
Qy      181 KERNGNHDPBMRKAVIYSKHEWSGDSSADKRYKGDPPAPRGVLDMGRDNI 240
Db      226 KERNGNHDPBMRKAVIYSKHEWSGDSSADKRYKGDPPAPRGVLDMGRDNI 285
Qy      241 PRSPTSPGEGFVNFYGMFGAQTADADKTWTHGNHYHAPNGLGMEVYSKFRNWE 300
Db      286 PRSPTSPGEGFVNFYGMFGAQTADADKTWTHGNHYHAPNGLGMEVYSKFRNWE 345
Qy      301 GYSDFDRGAYVITTPKSNATPADKYKQGW 331
Db      346 GYSDFDRGAYVITTPKSNATPADKYKQGW 376

RESULT 2
Q8KRJ2 ID Q8KRJ2 PRELIMINARY; PRT; 407 AA.
AC Q8KRJ2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transglutaminase precursor.
OS Streptomyces mobaraensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=35621;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO13819;
RA Kikuchi Y., Date M., Yokoyama K.-I., Umezawa Y., Matsui H.;
RT "Secretion of active form transglutaminase of Streptovorticillium
RT mobaraense in Corynebacterium glutamicum: Processing of pro-domain
RT with co-secreted subtilisin-like protease from Streptomyces
RT aldoagrisseolus."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF531437; AAM9595.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT SIGNAL.
SQ SEQUENCE 407 AA, 45684 MW, 10777FA04EAB2DF4 CRC64;

Query Match 100.0%; Score 1811; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.6e-128;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSDDRITPPAPPLDMPPRPSYGRAEVYNNYIRKQOYYSRDRKQOMTEQRRL 60
Db 77 DSDDRITPPAPPLDMPPRPSYGRAEVYNNYIRKQOYYSRDRKQOMTEQRRL 136
Qy 61 SYGCVGVTVWNSGQYFENRLAFASPEDRFKKEIKNGRPSGSTRAFEGRAVAKESFDEE 120
Db 137 SYGCVGVTVWNSGQYFENRLAFASPEDRFKKEIKNGRPSGSTRAFEGRAVAKESFDEE 196
Qy 121 KGFQARREVAIVMNALENHDESAVLDNLKKELANGDARLNEDARSPFYALNTPSF 180
Db 197 KGFQARREVAIVMNALENHDESAVLDNLKKELANGDARLNEDARSPFYALNTPSF 256
Qy 181 KERNGNHDPBMRKAVIYSKHEWSGDSSADKRYKGDPPAPRGVLDMGRDNI 240
Db 257 KERNGNHDPBMRKAVIYSKHEWSGDSSADKRYKGDPPAPRGVLDMGRDNI 316
Qy 241 PRSPTSPGEGFVNFYGMFGAQTADADKTWTHGNHYHAPNGLGMEVYSKFRNWE 300
Db 317 PRSPTSPGEGFVNFYGMFGAQTADADKTWTHGNHYHAPNGLGMEVYSKFRNWE 376
Qy 301 GYSDFDRGAYVITTPKSNATPADKYKQGW 331
Db 377 GYSDFDRGAYVITTPKSNATPADKYKQGW 407

RESULT 3

```

```

Q8KNY5 ID Q8KNY5 PRELIMINARY; PRT; 334 AA.
AC Q8KNY5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transglutaminase (fragment).
OS Streptomyces mobaraensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=35621;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27441;
RA Jiang S.-T., Tzeng S.-S., Wu W.-T., Chen G.-H.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY129279; AAN01353.1; -.
FT NON_TER
FT SIGNAL.
SQ SEQUENCE 334 AA, 38168 MW, 8C2356C48FC2CB35 CRC64;

Query Match 95.6%; Score 1732; DB 2; Length 334;
Best Local Similarity 93.7%; Pred. No. 1.1e-122;
Matches 310; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Qy 1 DSDDRITPPAPPLDMPPRPSYGRAEVYNNYIRKQOYYSRDRKQOMTEQRRL 60
Db 4 DSDDRITPPAPPLDMPPRPSYGRAEVYNNYIRKQOYYSRDRKQOMTEQRRL 63
Qy 61 SYGCVGVTVWNSGQYFENRLAFASPEDRFKKEIKNGRPSGSTRAFEGRAVAKESFDEE 120
Db 64 SYGCVGVTVWNSGQYFENRLAFASPEDRFKKEIKNGRPSGSTRAFEGRAVAKESFDEE 123
Qy 121 KGFQARREVAIVMNALENHDESAVLDNLKKELANGDARLNEDARSPFYALNTPSF 180
Db 124 KGFQARREVAIVMNALENHDESAVLDNLKKELANGDARLNEDARSPFYALNTPSF 183
Qy 181 KERNGNHDPBMRKAVIYSKHEWSGDSSADKRYKGDPPAPRGVLDMGRDNI 240
Db 184 KERNGNHDPBMRKAVIYSKHEWSGDSSADKRYKGDPPAPRGVLDMGRDNI 243
Qy 241 PRSPTSPGEGFVNFYGMFGAQTADADKTWTHGNHYHAPNGLGMEVYSKFRNWE 300
Db 244 PRSPTSPGEGFVNFYGMFGAQTADADKTWTHGNHYHAPNGLGMEVYSKFRNWE 303
Qy 301 GYSDFDRGAYVITTPKSNATPADKYKQGW 331
Db 304 GYSDFDRGAYVITTPKSNATPADKYKQGW 334

RESULT 4
Q8GR90 ID Q8GR90 PRELIMINARY; PRT; 416 AA.
AC Q8GR90;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transglutaminase.
OS Streptomyces cinamomeus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=33446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IPO 12852;
RA Yokoyama K.;
RT "Streptovorticillium cinamomeum IFO12852 Tgase gene."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB085698; BAC24766.1; -.
SQ SEQUENCE 416 AA, 46394 MW, 677DA9778829839D CRC64;

Query Match 82.1%; Score 1487; DB 2; Length 416;
Best Local Similarity 81.5%; Pred. No. 4.4e-104;

```

Matches 263; Conservative 26; Mismatches 35; Indels 0; Gaps 0;

QY 2 SDDRVTTPAEPLDMPDPYRPSYGRATVNNYIRKQOYSHRDRKQOMTEBQREWLS 61
 DB 87 SDDRTPPAEPDLDMPEAAYAYGGRATVNNYIRKQOYSHRDRKQOMTEBQREWLS 146
 QY 62 YGCVGYTVWNSGQYPTNRLAFASFPDEPRKXELKXGRPSGGEFRAEFGRVAKESFDEEK 121
 DB 147 YGCVGYTVWNSGQYPTNRLAFASFPDENKYKIDLKNTSPREDTEFRAEFGRVAKESFDEEK 206
 QY 122 GFORAREVAVNNALNADHESAYLDNLKKEIANGDALNEDASPFYSALNTSPFK 181
 DB 207 GFRARADVASVNNALNADHESAYLDNLKKEIANGDALNEDASPFYSALNTSPFK 266
 QY 182 BBNAGNDPSSAKAVIYSKHFWSGQDRSSADKRYGDPAPFAPGTGLVMSDRNP 241
 DB 267 BNDGNDPSSAKAVIYSKHFWSGQDRSSADKRYGDPAPFAPGTGLVMSDRNP 326
 QY 242 RSPFSPGEGFVNDYGFQAGTADADKTWTHGNHHAHNSGLAMHYESKPRNNSG 301
 DB 327 RSPAPFEGGVNFDYGFQAGTADADKTWTHGNHHAHNSGLAMHYESKPRNNSG 386
 QY 302 YSDFRGAYVITFTPKSMTAPADKYKQGW 331
 DB 387 YADFDRGAYVITFTPKSMTAPADKYKQGW 416

RESULT 5

Q9R1S1 PRELIMINARY; PRT; 411 AA.

AC Q9R1S1; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Transglutaminase precursor (EC 2.3.2.13).
 GN TMASE
 OS Streptomyces cinamomeus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=53446;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 683.68;
 EX MEDLINE=98336622; PubMed=9672751;
 RA Duran R., Juncua M., Schmitter J.M., Gancet C., Goulas P.,
 RT "Purification, characterization, and gene cloning of transglutaminase
 (Tgase) from Streptococcillus cinamomeum CBS 683.68."
 RL Biochimie 80:313-319 (1998).
 DR EMBL: Y08820; CAA70055.1; -
 KM Acyltransferase; Signal; Transferase.
 FT SIGNAL
 FT CHAIN 82 411 POTENTIAL.
 SQ SEQUENCE 411 AA; 46499 MW; 01A0CA2EP4C388B CRC64;

Query Match 79.1%; Score 1432; DB 2; Length 411;
 Best Local Similarity 79.7%; Pred. No. 6, 1e-100;
 Matches 263; Conservative 27; Mismatches 36; Indels 4; Gaps 3;

QY 2 SDDRVTTPAEPLDMPDPYRPSYGRATVNNYIRKQOYSHRDRKQOMTEBQREWLS 61
 DB 86 SDDRTPPAEPDLDMPEAAYAYGGRATVNNYIRKQOYSHRDRKQOMTEBQREWLS 145
 QY 62 YGCVGYTVWNSGQYPTNRLAFASFPDEPRKXELKXGRPSGGEFRAEFGRVAKESFDEEK 121
 DB 146 YGCVGYTVWNSGQYPTNRLAFASFPDENKYKIDLKNTSPREDTEFRAEFGRVAKESFDEEK 204
 QY 122 GFORAREVAVNNALNADHESAYLDNLKKEIANGDALNEDASPFYSALNTSPFK 181
 DB 205 GFRARADVASVNNALNADHESAYLDNLKKEIANGDALNEDASPFYSALNTSPFK 263
 QY 182 BBNAGNDPSSAKAVIYSKHFWSGQDRSSADKRYGDPAPFAPGTGLVMSDRNP 241
 DB 264 BNDGNDPSSAKAVIYSKHFWSGQDRSSADKRYGDPAPFAPGTGLVMSDRNP 323

QY 242 RSPFSPGEGFVNDYGFQAGTADADKTWTHGNHHAHNSGLAMHYESKPRNNSG 301
 DB 324 RSPAPFEGGVNFDYGFQAGTADADKTWTHGNHHAHNSGLAMHYESKPRNNSG 383
 QY 302 YSDFRGAYVITFTPKSMTAPADKYKQGW 331
 DB 384 YADFDRGAYVITFTPKSMTAPADKYKQGW 411

RESULT 6

Q8U0G0 PRELIMINARY; PRT; 737 AA.

AC Q8U0G0; 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Capsid protein.
 OS Adeno-associated virus 7.
 OC Viruses; ssDNA viruses; Parvoviridae; Dependovirus.
 OX NCBI_TaxID=202812;
 RN 11
 RP SEQUENCE FROM N.A.
 RX PubMed=12192090;
 RA Gao G.P., Alviria M.R., Wang L., Calcedo R., Johnston J., Wilson J.M.;
 RT "Novel adeno-associated viruses from rhesus monkeys as vectors for
 human gene therapy."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859 (2002).
 DR EMBL: AF513851; AAN03855.1; -
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 SQ SEQUENCE 737 AA; 81652 MW; D63BCE7583140D0 CRC64;

Query Match 6.7%; Score 122; DB 12; Length 737;
 Best Local Similarity 22.8%; Pred. No. 0.81;
 Matches 76; Conservative 45; Mismatches 161; Indels 52; Gaps 16;

QY 18 DYPSPYGRATVNNYIRKQOYSHRDRKQOMTEBQREWLSYGVGYTVWNSGQY-- 75
 DB 4 DGYLPDW--LEDNLSEGIREW--WDLKPGAPKPKANQKQDNGSLVLPQYKYLGFNG 58
 QY 76 -----PNNRLAFSPDEDR-FNDELKNG-RP--RSGETAAEFEGVAESDEDEKQORA 126
 DB 59 LDKGEPVNAADAALEHDPAYDQKAGDNVYLYNHADHOFERLQ--DTSFGNLG 115
 QY 127 REVASVNNRLENAHDESAVLDNLKKEIANGDALNEDASPFYSALNTSPSKERNNG 186
 DB 116 RAVFOAKKRVLEPL-----GLVESGAKTAPAKKRPVBPQSPSPSIGIKK---- 163
 QY 187 NHDSEKKAIVISKHFWSGQDRSSADKRYGDPAPFAPGTGLVMS-----RDKNIP 241
 DB 164 GQOQAPARRL-----NFGQTGDESVDPQPLGEPPAPSPSGTVAAGGAPVADNNBG 218
 QY 242 RSPFSPGEGFVNDYGFQAGTADADKTW--TGNNHYHAPNSGLAMHYESKPRN 298
 DB 219 ADGCVNNSGWHDCSTWLCDRVITTSRT WALPTVNNHLYQIISSEFAGSTNDNTTFGY 277
 QY 299 SE--GYSDPDRGAYVITFTPKSMTAPADKYKQGW 330
 DB 278 STPWGYDFDNR--FCHFSFSPRWQRL--INNWW 306

RESULT 7

Q17240 PRELIMINARY; PRT; 1217 AA.

AC Q17240; 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Sericin1B.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

OC Bombycidae, Bombyx.
 OX NCBI_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=European 200x300; TISSUE=Middle silk gland;
 EX MEDLINE=97382906; PubMed=9219370;
 RA Garel A.A., Deleage G.G., Prudhomme J.J.,
 RT "Structure and organization of the Bombyx mori sericin 1 gene and of
 the Sericini deduced from the sequence of the Serib cDNA";
 RL Insect Biochem. Mol. Biol. 27:469-477(1997).
 DR EMBL: Z48802; CA88741.1;
 SQ SEQUENCE 1217 AA; 123411 MW; 4C5789F875E866D0 CRC64;

Query Match 6.3%; Score 114.5; DB 5; Length 1217;
 Best Local Similarity 20.4%; Pred. No. 5.8;
 Matches 60; Conservative 47; Mismatches 136; Indels 51; Gaps 11;

QY 33 NTRKQYVYSHRDKQKQTEBQ---EMLSYGVGVTVTVNSQ-----YPTNR 79
 DB 124 NYVSDQAVAAASDABENRSDQANWMDGSGVYVSDRSGASRRARQANYSDXD 183
 QY 80 LAFASPEDREFKNEIKGPRSGEPTAEFEGRVAKESPEEKGFQARAVAYVMRALEN 139
 DB 184 ITPASDSDSRADSSRRN-----AYNRDSD-----GSBSAGLSDBSASS 223
 QY 140 AHDESAIYDNLKKELANGNDALRNEDARSPFYALRNTPSFEKNGNDPDS---RMK 194
 DB 224 SKNDNFVYRTYDST--GGQAKSSRSBSHSDSDAYNNSPDGYNAGTRSSSTNKKKAS 281
 QY 195 AVTYS--KTFMSGGDRSSADKRYGDPDAPFAPAGTGLVMSMDRINPSPSPGEGFV 252
 DB 282 STIYADKQDIRAANDRSSKOLKSSAIOISGP-KATSV--SSDRORYSNDKKSXDAYV 338
 QY 253 NPFYGMFG--AQTEADADKTVWNGHVAHPAGSLGAAHVYSEKFRWMSGYSD 304
 DB 339 GRD---GTVAISNKDSKTSRQSNNTYADQNSVSDASADQTSKYDRGISD 388

RESULT 8
 ID Q8QJF8 PRELIMINARY: PRT; 738 AA.
 AC Q8QJF8;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Capsid protein.
 OS Adeno-associated virus 8.
 OC Vituuses; ssDNA vituuses; Parvoviridae; Parvovirinae; Dependovirus.
 OX NCBI_TaxID=202813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12192090;
 RT "Novel adeno-associated vituuses from rhesus monkeys as vectors for
 human gene therapy";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:11854-11859(2002).
 DR EMBL: AF513852; JAM03857.1;
 DR InterPro: IPR001403; Parvovirinae.
 DR Pfam: PF00740; Parvovirinae.
 SQ SEQUENCE 738 AA; 81756 MW; 59BC2BE1B48AA14 CRC64;

Query Match 6.3%; Score 113.5; DB 12; Length 738;
 Best Local Similarity 22.9%; Pred. No. 3.5;
 Matches 78; Conservative 46; Mismatches 153; Indels 63; Gaps 20;

QY 18 DVPSPSGAETVANNVTRKQYVSHRDKQKQTEBQREMLSYGVGVTVTVNSQY-- 75
 DB 4 DGLPDM--LEDNUSGGRK--WALKPAPAPKANOQKQDDRGVLDPGYRLSPFPG 58
 QY 76 -----PTNRLAFASDEDR-FKNEKNG-RP--RSGETRAEFGRVAKESFDEKGFQRA 126
 DB 59 LDKGEPVNAADAAALHDKAVDQQLQAGNDPFLRVNVAADAFQGRLE--DISFGNIG 115

QY 127 REVASVNRALENAHDESAVLDNLKKELANGNDALRNEDARSPFYALRNTPSFEKNG 186
 DB 116 RAVFOAKKRVLEPL-----GLVEGAKTAPKPKPBPSPQSPDSSTGIGK----- 163
 QY 187 NNDPBRMAVYVSHFMSGQDSSADKRXKGDADARAP-GTGLVMSMDRINPSPSPG 245
 DB 164 GOQPARKRL-----NFGQGTDESVDPDQFGEPPA--APSGVGNTPAAGGAPVADN 215
 QY 246 SPG-----EGFVNFPGVFGAQTADADKTVW---THGNHYH--APNGSIGM--HV 290
 DB 216 NRGADGVSSSGMNHCDSTGLDRTYTTSTRL-WALFTYNNHLYKQISNGTGGATINDT 274
 QY 291 YSKFRNWSGYSDFDGAIVITFIPKSNWTPAPKXQGV 330
 DB 275 YFGYSTPW--GYDPNR--FHCHFSRDMQRL--INNNV 307

RESULT 9
 ID Q9W2X8 PRELIMINARY: PRT; 1161 AA.
 AC Q9W2X8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE CG15311 protein.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktargolu L., Beasley E.M.,
 RA Beeson K.Y., Besos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokya D., Botchan M.R., Bouck J., Brokstein P., Bottrick P.,
 RA Borkov D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Idagala M., Kalush F., Karpen G.H., Ke Z., Kemtson U.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laekko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclio J.M.,
 RA Palazolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA She B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasman D.A., Weinstein G.K., Wu D., Yang S., Yao Q.A.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).

DR EMBL: AE003449; AA046559.1; -
 DR FLYBASE; FBgn0030182; CG15311.
 SQ SEQUENCE 1161 AA; 130214 MW; DD01D5B40805ID36 CRC64;

Query Match 6.2%; Score 112.5; DB 5; Length 1161;
 Best Local Similarity 19.1%; Pred. No. 7.7;
 Matches 67; Conservative 66; Mismatches 152; Indels 65; Gaps 13;

QY 8 PPAELRMDPDPYRPSYGRATVNNYIRKQVYSHDRKQMTBEOREMLSYGCVG 67
 B 180 PQLERBQVEHEENAFGR-QSFAYKKAKMHEGQSQDKKDDGDE----- 226
 QY 66 TWVNSGYPTNRLAFASFPEDRFKNELKNGP--RSGETRAEFGR-----VAKESFDE 119
 B 227 VDLIDGATPSPQL-----FPAIILNERKFNKPSAQSTGKRPQLCKKDKQKQKQVQ 282
 QY 120 EKGFGARAEVAYVNNALNADBSAYLDNLKELANGNDALNEDARSPYSALRTPS 179
 B 283 EDEQKHSQVDEAIGDIEDLDNDILPTNNGEDDDGDADIDDD--EDTKSALDDEL 340
 QY 180 FKERNGGNDPSPKAKVIVYSKHFWSGQDRSSSADKRYGDPDAFRPAPGTGLVMSRDN 239
 B 341 AKKYVATSTTKVPTTLATSKTISRSSSSTTTTMAISTATSPSPSTTTK----- 394
 QY 240 IPRSTSPGSGS-----FVAFDYGMFGAQTADADKTYWTHGNHYAPN 282
 B 395 -PRPT-PTIGRLKKNFGLSQPTRYGSDYDGRIRAQTD-DTT-----BEKMD 444
 QY 283 GSLGAMVYESKFRNWSGVSDFD---RGAYVTFPRKSNWNPAPKVKYQ 328
 B 445 ENVEGLHYDTSGSSSRKLVSPDEKSENYLSTTFGMM-ATKCKQ 493

RESULT 10

ID 081525 PRELIMINARY; PRT; 758 AA.
 AC 081525;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE B1103C09.11 protein (P0451D05.19 protein).
 GN B1103C09.11 OR P0451D05.19.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriatridae; Oryzaceae; Oryza.
 CX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsunoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC clone:B1103C09."
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsunoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone:P0451D05."
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AP003333; BAB91821.1; -
 DR EMBL; AP003253; BAB92317.1; -
 DR Gramene; O81525; -
 SQ SEQUENCE 758 AA; 81761 MW; 2A95E763198FE12C CRC64;

Query Match 6.2%; Score 111.5; DB 10; Length 758;
 Best Local Similarity 22.0%; Pred. No. 5.2; Indels 91; Gaps 14;
 Matches 78; Conservative 40; Mismatches 146;

QY 3 DDRVPPAELRMDPDPYRPSYGRATV--NNYIRKQO-----VYSHD----- 46
 B 125 DGRAPPRVAVAGGEGEKGVGEAAVVEDGGEKKEQEPFAVPTSGAFYVHNDRPFGE 164

QY 47 --GRKQMTBEOREMLSYGCVGVTWVNSGYPTNRLAFASFPEDRFKNELKNGP 104
 B 185 SRGRRRMRGKRLVDAD--DQAMVHD-RFEEMNLHEHNYEDKMSGRRRGGGGRT 241
 QY 105 RAPEEG-----RVAKESDEBKGRQARREY 129
 B 242 RGTGFGARQKGYGNDINNHNQRPQVVRGRRRYEAAKNND-VVGFQRKQPA 300
 QY 130 ASVNNRNLNADBSAYLDNLKELANGNDALNEDARSPYSALRTPSP----- 180
 B 301 RSRSAASASVRESGQTLNASEMAPPKKNVNVSSLSNASPFYPSGASNPDFSVAQR 360
 QY 181 --KERNGND--PSRKAIVYSKHFWSGQDRSSSADKRYGDPDAFRPAPGTGLVMSR 236
 B 361 RDNMQAGSKRVPPSGMMDNNAK-----VQSGPAVRDYGARDRFQHALGP----- 407
 QY 237 DRNTPRSPSPGEGVNPDYGMFGAQTADADKTY--WTHGNHYAPNGSLGAMH 289
 B 408 --VRQSPRSG--TSLSNGFAASTVNHGSSVVRTOGGNGIPSNQSTSLH 456

RESULT 11

ID 065311 PRELIMINARY; PRT; 736 AA.
 AC 065311;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Capsid protein.
 OS Adeno-associated virus 3.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 CX NCBI_TaxID=46350;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H.
 RX MEDLINE=96266430; PubMed=8661429.
 RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
 RT "Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3."
 RL Virology 221:208-217(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H;
 RA Muramatsu S., Brown K.E.;
 RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL; U48704; AAC55049.1; -
 DR InterPro; IPR001403; Parvo Coat.
 DR Pfam; PF00740; Parvo Coat; 1.
 SQ SEQUENCE 736 AA; 81660 MW; AFPIEF47B5C67A10 CRC64;

Query Match 6.1%; Score 110; DB 12; Length 736;
 Best Local Similarity 22.4%; Pred. No. 6.5;
 Matches 77; Conservative 49; Mismatches 144; Indels 74; Gaps 22;

QY 18 DVPYPSYGRATVNNYIRKQVYSHDRKQMTBEOREMLSYGCV--GYTWNSGQ- 74
 B 4 DGYLPDW--LEDNLISGIFEW--WALKPGVPGKAKQCHONRGLVLPQKYLGGNG 58
 QY 75 ---YPTNRLAFASFPEDRFKNELKNGP--RSGETRAEFGRVAKESFDEBKGR 126
 B 59 LDKGEPVNEADALHDAVDQQLKGDNLYLKNHADBFQERQJ--DTSGFNGLG 115
 QY 127 REVASVNNRNLNADBSAYLDNLKELANGNDALNEDARSPYSALRTPSPFKERN 184
 B 116 RAVQAKRILTEPLGVEEAATAKPAKK-----GAVDQSPQEPDSS 156
 QY 185 GG---NHDSRKAVIVYSKHFWSGQDRSSSADKRYGDPDAFRPAP-GTGYVMSRDN 239
 B 157 SGVSGSKQPARKRL-----NFGQTGSESVYDQPAAGEPPA---APTSLGNTWASGG 208
 QY 240 IPRSTSPG-----EGFVNPDYGMFGAQTADADKTY--TGNH-YAAPNGSLGAM 288

Db 209 APMADNNBAGDGVSSGNMHCDSQMLGDRVITTTTFRF WALPYNNHLYKQISSQSGAS 267
 QY 289 HYESKFRNMSE--GYSDPDGRGAVYTFIPKSWNTAPDXKQGW 330
 Db 268 N--DNHFYFGVSTPMGVFPDNR--FHCFSRDMQRL--INNWN 304

RESULT 12
 ID 007351 PRELIMINARY; PRT; 919 AA.
 AC 007351.
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Protease precursor.
 GN PRVY.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=017;
 RX MEDLINE=89013889; PubMed=3050359;
 RA Alm R.A., Stroeber U.H., Manning P.A.;
 RT "Extracellular proteins of *Vibrio cholerae*: Nucleotide sequence of the
 RT structural gene (hlyA) for the haemolysin of the haemolytic EL Tor
 RT strain 017 and characterization of the hlyA mutation in the non-
 RT haemolytic classical strain 569B."
 RL Mol. Microbiol. 2:481-488(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=017;
 RA Manning P.A.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y00557; CAA6636.1; .
 DR MEROPS; M06.002; .
 DR InterPro; IPR000601; PKD domain.
 DR InterPro; IPR006025; Zn_MTPeptide.
 DR Pfam; PF00801; PKD; 2.
 DR SMART; SM00089; PKD; 2.
 DR PROSITE; PS50093; PKD; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR Signal.
 FT SIGNAL.
 SQ SEQUENCE 919 AA; 101996 MW; 6A8074801FBB8D CRC64;

Query Match 6.0%; Score 109; DB 2; Length 919;

Best Local Similarity 21.8%; Pred. No. 10; Indels 78; Gaps 16;
 Matches 67; Conservative 52; Mismatches 111;

QY 42 YSHR-DGRKQMTTEQREWLSTYGVGTWNSQYPTNRL-AFASFDED-RFKNELKNGR 98
 Db 446 YSNRGDLKRMNRP-----LTI PASQATLRFKWPQLKDYANVLNKG 493
 QY 99 PRSGETAEFE---GRVAKSFDEKGFQPAR-EVASVMDALNAD-----ES 144
 Db 494 PIAGNITMDPDKSLVPAISQSGDWDAQFDISAWAGQVELADYLTQGLAMG 553
 QY 145 AYLDNLEKELANGNDALRNEDARSPYSALRNPSRK-ERNGNHDSSKKAVTSKRP 202
 Db 554 LYVDLRLLEVDGNQTLIDNASTSF-----AQGFTKNGGHE-----ANNH 596
 QY 203 WSGQDRSSS-----ADKKKYGDPPAPFAPGTGLVMSRDNRNIPRPTSGEGFVNFY 256
 Db 597 YLLQWRSHNDVDCGLANLRKFGQLMSPFGILVWYVDESVALDNWVGK--HFGG----- 647
 QY 257 GMPGAQTADADKTVTTHGNTHAPNSGLAMVYSKPRMNSBGYSFPGAGAVITFPD 316
 Db 648 GMLG-RVADADQNALVMSKTG-----EVAQTRFQVADATFSLFDQAP--LKLTV 692
 QY 317 KSWNTAPD 324

Db 693 ADGNTLED 700

RESULT 13
 ID 096C72 PRELIMINARY; PRT; 401 AA.
 AC 096C72;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC014625; AAH14625.1; .
 KW Hypothetical protein.
 FT NON-TER
 SQ SEQUENCE 401 AA; 48615 MW; D9ADPDC029A0851D CRC64;

Query Match 6.0%; Score 108; DB 4; Length 401;
 Best Local Similarity 21.5%; Pred. No. 4.1;
 Matches 68; Conservative 33; Mismatches 118; Indels 98; Gaps 12;

QY 39 QQYSHRDGRKQMTTEQREWLSTYGVGTWNSQYPTNRLAASFDKPFKELKNGR 98
 Db 7 EERNHLEERKQREKERR-----ITYYREKEEEOORAAEOMLKEREREERARA 57
 QY 99 PRSGETAEFEGRVAKSFDEKGFQPAR-EVASVMDALNADBSAYLDNLEKELANGN 158
 Db 58 KRESELR-EYQERVAKLEBEVERKQRQELIEBERRR-----EZEKRLGD 102
 QY 159 DALRNDAE-----SPFYSALRNTPSEKE-RANGNDPSRMKAVIYSKPFMS 204
 Db 103 SLSLRDSWGRDSEGTWRKGFADENRGRGPEKWRNRGRGD----- 147
 QY 205 GQDRSSSADK---RKYGPD---AFRP---APGTGLVMSRDNRNIPRPTSGEGFVN 253
 Db 148 -EDRSRRRDESRRLGDDEDEPSLFPDDRVPRGMDDGRGPRGEGEDRSRGADD 206
 QY 254 FDYGMFGAQT-----ADADKTVTTHGNTHAPNSGLAMVYSKPRMNSBGYSDFDR 307
 Db 207 DRSNNTDDDRPPRIIDEDRGKWRHADDPRFRGL-----DDBR 248
 QY 308 GAVYTFIPKSWNTAPD 324
 Db 249 G-----SWRTADE 256
 RESULT 14
 ID 09H0G5 PRELIMINARY; PRT; 558 AA.
 AC 09H0G5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN DKF2P434K1421.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Niemann S., Weill B., Wellenreuther R., Gassenhuber J., Glasel S.,
 Ansoerge W., Boecher H., Bauersachs S., Blum H.,

RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
 RA Mewes H.W., Ockenwaelder B., Obermayer B., Tampe J., Heubner D.,
 RA Mamuth R., Korn B., Klein M., Poustka A.,
 RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
 RT Analysis of 500 Novel Human Complete Protein Coding Human cDNAs."
 RL Genome Res. 11:422-435 (2001).
 DR EMBL: AL136806; CAB6740.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 558 AA; 66390 MW; 99B7BDCFD06F39D CRC64;

Query Match 6.0%; Score 108; DB 4; Length 558;
 Best Local Similarity 23.8%; Pred. No. 6.3;
 Matches 59; Conservative 31; Mismatches 86; Indels 72; Gaps 11;

QY 36 KRWQVYSHRDKQMTSECEWLSYGVVWNSGQYPTNRLAFASFDREKFKELK 95
 DB RDQENHYTDYKEDSHRHE-----ASHDSHWKHEG 361
 QY 96 NGRRSGETAEFEGVAEAFDEKGFQAPR--VASVNRPALENADH--SAYLDN 149
 DB 362 EDKFRARDQR-ERSDREVMREKREKYSQREEDRQNDQNRPSKKEKSKAKEEH 420
 QY 150 LK--KELANGNDALRNEEDAR-----SPFYSA-----LRETPS 179
 DB 421 MKVAKERYENDKYRPREKREYGVOSERRNDRKSSSPNSRAKDKFLDQERSNRRMAK 480
 QY 180 FKERNGNHDPSPMKAVIYKHPWS--GQDRSSASAKRYKDPDA--FRPAPGTGLVDM 234
 DB 481 DXERN--QEKPSNSESLGAKHRLTEBQEKKEQGER---PPEAVSKFAKRNNEETVMS 534
 QY 235 SRDKNIPR 242
 DB 535 ARDYLAR 542

RESULT 15
 ID 056139 PRELIMINARY; PRT; 736 AA.
 AC 056139;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DS Capsid Protein VP1.
 CS Adeno-associated virus 3B.
 CC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 CX NCBI_TaxID=68742;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98080418; PubMed=9420229;
 RA Rutledge E.A., Halbert C.L., Russell D.W.,
 RT "Infectious clones and vectors derived from adeno-associated virus
 RT (AAV) serotypes other than AAV type 2.";
 RL J. Virol. 72:309-319 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rutledge E.A., Russell D.W.,
 RT Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF028705; AAB95452.1; -.
 DR InterPro: IPR001403; Parvo_coat.
 DR Pfam: PF00740; Parvo_coat; 1.
 SQ SEQUENCE 736 AA; 81906 MW; D52331AD5FD70F CRC64;

Query Match 5.9%; Score 107; DB 12; Length 736;
 Best Local Similarity 22.5%; Pred. No. 11;
 Matches 76; Conservative 47; Mismatches 153; Indels 62; Gaps 21;

QY 18 DPKPSYGRATVNNYIRKQVYSHRDKQMTSECEWLSYGVV--GYVWNSGQ- 74
 DB 4 DGYLPDW--LEDNISGIREW---WALKPVGPQKANOQHONRRGLVLPQKYLGPNG 58
 QY 75 ----YPTNRLAFASFDREK-FKDELKNG-RP--RSGETAEFEGVAEAFDEKGFQRA 126
 DB 59 LDKGEFVNEADALAHDKAIDQQLKAGNDYILKINHADAEPQERLQF---DTSGGNLG 115

QY 127 REVAVVNRALZNAHDSAYLDNIKELANGNDALRNEEDARSPFYSALENTPSEKERNCG 186
 DB 116 RAVFQAKKRILBPL-----GLVEAAKTAPGKKRPVDSF-----QEDSSSGVGKS 162
 QY 187 NHDSSMKAVIYKHPWSGQDRSSADKRYKDPDAFRAP-GTGLVDMSPDRNIPPSPT 245
 DB 163 GKQFARKRL----NFGQTDGDSBSVPDPQPLGPPA---APTSLGNTTMAOGGAPVADN 214
 QY 246 SPG-----EGFVNPDYGMFGAQTEADADKTVM---THGNH-YHAPGSLGAMHYESK 294
 DB 215 NEGADGVNSSGNWHDCSOMLGRVITTTSTRT-WALPTNNHLYQISSGGSASN--DNH 271
 QY 295 FRNWSR--GYSDPDRGAYVITFIPKSNNTAPDKVKQGW 330
 DB 272 YFGYSTPWGYPDENR--FCHHSPRDMQRL---INNWN 304

Search completed: September 22, 2003, 12:45:04
 Job time : 44 secs